Single-Molecule FRET Guided Modeling of RNA Structure and Dynamics

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde (Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Fabio Steffen

von

Brütten ZH

Promotionskommission

Prof. Dr. Roland K. O. Sigel (Vorsitz)

Prof. Dr. Benjamin Schuler

Prof. Dr. Helmut Grubmüller

Prof. Dr. Richard Börner

Zürich, 2020

"Everything that living things do, can be understood in terms of the jigglings and wigglings of atoms"

The Feynman Lectures on Physics, Volume I [1]

Table of Contents

Publications

- **FRET-assisted modeling of riboswitch structural dynamics**
  F. D. Steffen, S. Gallo, R. K. O. Sigel‡, R. Börner‡, *in preparation*, **2020**.

  contribution: FS designed the research together with RB, recorded and analyzed single-molecule FRET data, wrote analysis code for modeling and simulations and drafted the manuscript.
  ⟶ Chapter 5

- **Metal ions and sugar puckering balance single-molecule kinetic heterogeneity in RNA and DNA tertiary contacts**
  F. D. Steffen*, M. Khier*, D. Kowerko, R. A. Cunha, R. Börner‡, R. K. O. Sigel‡, *Nature Communications* **2020**.
  doi.org/10.1038/s41467-019-13683-4

  contribution: FS analyzed the single-molecule data, set up and analyzed the MD simulations and wrote the paper together with RB.
  ⟶ Chapter 4

- **Stick, flick, click – DNA guided fluorescent labeling of long RNA for single-molecule FRET**
  F. D. Steffen‡, R. Börner, E. Freisinger, R. K. O. Sigel‡, *Chimia* **2019**.
  doi.org/10.2533/chimia.2019.257

  contribution: FS wrote the review.
  ⟶ Chapter 3

- **Site-specific two-color labeling of long RNAs for single-molecule FRET**
  M. Zhao, F. D. Steffen, R. Börner, M. F. Schaffer, R. K. O. Sigel‡, E. Freisinger‡, *Nucleic Acids Research* **2018**, *46*, e13.
  doi.org/10.1093/nar/gkx1100

  contribution: FS measured and analyzed the photophysics of the dye-labeled RNA and wrote the paper together with MZ.

- **An atomistic view on carbocyanine photophysics in the realm of RNA**
  F. D. Steffen, R. K. O. Sigel, R. Börner‡, *Physical Chemistry Chemical Physics* **2016**, *18*, 29045–29055, Front Cover.
  doi.org/10.1039/c6cp04277e

  contribution: FS measured and analyzed the photophysics, set up and analyzed the MD simulations, designed the cover and wrote the paper together with RB. The paper is based on Master thesis work carried out by FS at UZH and is therefore not included as a separate chapter in this thesis.

* authors contributed equally

‡ corresponding author

Open Access

Chapters 3 and 4 of this thesis are published in peer-reviewed journals under an open access model. They are reprinted here in the form of the author accepted manuscript (AAM) prior to typesetting. The content of Chapter 3 is published in *Chimia* under the Creative Commons Attibution-NonCommercial 4.0 (CC BY-NC-ND 4.0). The content of Chapter 4 is published in *Nature Communications* and is licensed under the Creative Commons Attribution 4.0 (CC BY 4.0).

Source code developed in this PhD thesis is available on Github at github.com/fdsteffen and github.com/RNA-FRETools.

Abstract

Dynamics are central to the function of biomolecules. In the field of RNA, riboswitches are a prime example where regulatory function is encoded by structural transitions. In this work, we used single-molecule fluorescence spectroscopy together with molecular simulations to probe such conformational dynamics. The first part of this thesis reviews established and novel labeling approaches to site-specifically tag nucleic acids with fluorescent markers for Förster resonance energy transfer (FRET) applications. We characterize bioconjugated dyes in terms of their photophysics and introduce computational tools that help in selecting informative distance coordinates.

Biologically active RNA molecules are composed of recurrent, well-conserved modules connecting secondary and tertiary structure. Their systematic annotation over several decades led to the notion that RNA folding can be understood by the thermodynamics and kinetics of the constituting building blocks. Here, we chose a long-range tertiary contact reaching from the core of a group II intron to its flanking 5′-exon. The structure of the isolated contact was previously solved by NMR and allowed us to link chemical features of the ribose backbone and metal ion coordination with dissociation rates measured by single-molecule FRET. We speculate that kinetic heterogeneity in exon recognition has important implications on ribozyme catalysis.

Finally, we turn to a coenzyme B_{12} riboswitch whose mechanism has remained elusive owing to its structural complexity. Based on the consensus sequence and fragments of other cobalamin riboswitches we built a homology model of the *E. coli btuB* RNA and probed its dynamics by single-molecule FRET. We found a Mg^{2+} dependent conformational equilibrium which is thought to coordinate folding of the metabolite binding aptamer with the peripheral expression platform.

Kurzfassung

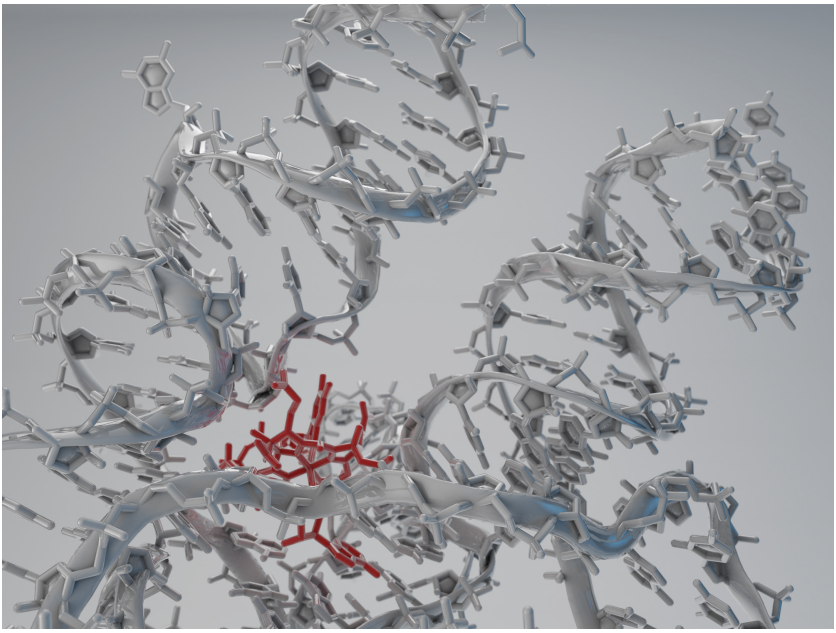
Biomoleküle sind ständig in Bewegung. Als Riboschalter interagieren Ribonukleinsäuren mit verschiedenen Metaboliten in der Zelle. Die mit der Bindung einhergehenden strukturellen Veränderungen der RNS steuern die Expression von Genen, die an der Biosynthese und dem Transport ebenjener Metaboliten beteiligt sind. Daraus ergibt sich ein Rückkopplungsmechanismus auf RNS Ebene, wie wir ihn aus Signalkaskaden von Proteinen kennen. Diese Arbeit befasst sich mit solchen Konformationsänderungen, welche wir mittels Einzelmolekülspektroskopie und Moleküldynamik (MD) Simulationen untersuchen. Der erste Teil gibt einen Überblick über bestehende und neu entwickelte Markierungsverfahren, deren Ziel es ist, einen fluoreszenten Farbstoff präzise in eine RNS einzubauen und diese dadurch spektroskopisch sichtbar zu machen. Mittels Fluoreszenzresonanzenergietransfer (FRET) zwischen zwei Farbstoffen können Biomoleküle im Nanobereich vermessen werden. Die Photophysik erlaubt derweil Rückschlüsse auf die chemische Umgebung der Farbstoffe. Wir haben computergestützte Methoden entwickelt, um diese Effekte zu quantifizieren und funktionell informative Reaktionskoordinaten zu wählen.

Biologisch aktive RNS Moleküle sind hierarchisch gegliedert und bestehen aus strukturell konservierten Modulen, welche die helikalen Elemente zu einer Tertiärstruktur verknüpfen. Im zweiten Teil der Arbeit beschäftigen wir uns mit einem Tertiärkontakt, der eine wichtige Rolle bei der Erkennung von Exons im Reifungsprozess der RNS spielt. Das Motiv ist Teil eines Introns und bindet eine komplementäre Sequenz auf dem flankierenden 5′-Exon. Dessen Bindungskinetik haben wir auf Einzelmolekülebene bestimmt und dabei abhängig von der Art des Zuckers heterogene als auch homogene Dissoziationsraten gefunden. Mit einer Kombination aus Magnetspinresonanz (NMR) und MD Simulationen konnten wir Metallbindungsstellen identifizieren sowie Unterschiede in der Konformationsdynamik von Ribose und Desoxyribose nachweisen und diese mit der kinetischen Heterogenität in Verbindung setzen. Mögliche Auswirkungen auf die Katalyse von Ribozymen haben wir am Beispiel von selbst-spleissenden Gruppe II Introns aufgezeigt.

Zum Schluss betrachten wir einen Vertreter der eingangs erwähnten Riboschaltern. Der molekulare Mechanismus dieser B_{12} abhängige RNS konnte bislang aufgrund der strukturellen Komplexität nicht abschliessend geklärt werden. Auf Basis von homologen Sequenzen sowie Fragmenten bestehender 3D-Strukturen haben wir ein Computermodel erstellt und dieses mittels FRET auf Dynamik hin untersucht. Dabei fanden wir ein Magnesium-abhängiges Gleichgewicht, welches möglicherweise die Faltung des metabolitbindenden Aptamers und der peripheren Expressionplattform koordiniert.

Chapter 1

A brief history of nucleic acids



Inside a B_{12} riboswitch

A regulatory RNA acts as a genetic switch to turn off gene expression in response to binding of the coenzyme adenosylcobalamin.

1.1 Genes split in pieces

In 1977, Phil Sharp and Richard Roberts contemporaneously discovered that eukaryotic genes are not contiguous but split into pieces.[2,3] While examining restriction fragments of a human adenovirus deoxyribonucleic acid (DNA) under an electron microscope, they noticed that the genomic DNA was interrupted by segments that do no longer appear in the mature messenger RNA (mRNA) (Fig. 1.1a). This observation suggested that the transcribed mRNA precursor is separated into multiple protein-coding sections (exons), interspersed by non-coding elements which we nowadays call introns. The introns need to be removed (or "spliced") from the primary transcript before the exons can be joined together to form the mature mRNA. Today, we know that most eukaryotic genes contain multiple exons that are spliced in alternative ways to produce various protein isoforms. By breaking the "one gene, one polypeptide" rule proposed by Beadle and Tatum in their 1941 seminal paper[4], alternative splicing has greatly expanded the proteome in higher eukaryotes.[5] Recombination of introns can reshuffle exons from one location in the genome to another. In some cases, the exons are brought into a new transcription unit to produce new protein assemblies. In this way, intron recombination and associated exon reshuffling generates protein diversity and assorts new functions in an organism.[6]

1.2 Ribozymes – relicts of an RNA world?

At the time of Sharp and Roberts' discovery, RNA was primarily thought of as an intermediary that ferries genetic information from DNA to proteins.[17,18] It was only a few years later, in early 1983, when the labs of Thomas Cech and Sidney Altman independently showed that RNA is able to self-splice from a plasmid transcript.[19,20] Their finding that certain classes of RNA, called ribozymes, can catalyze chemical reactions much like their protein counterparts, lent strong support to the so-called RNA world hypothesis.[21] It postulates that life on Earth originated from a self-replicating polymer capable of catalysis. One key puzzle piece to the existence of an RNA world is the pre-biotic synthesis of ribonucleosides.[22] Recently, a plausible geochemical scenario has been proposed that unifies the purine and pyrimidine synthetic routes under primordial conditions.[8–10] According to this model, RNA building blocks could have been produced from just a few starting compounds

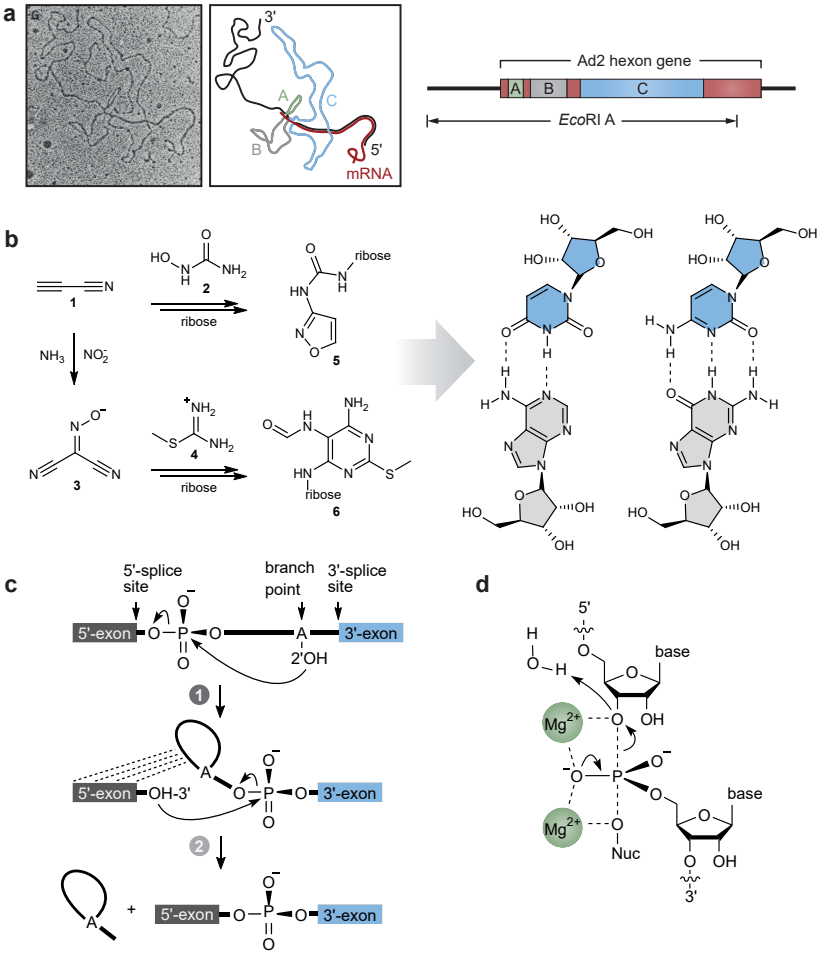


Fig. 1.1 | **The discovery of RNA splicing and the hypothesis of an RNA world.** **a** (*left*) electron micrograph (reproduced from ref. [2]) and interpreted diagram of the coding strand of adenoviral (Ad2) *Eco*RI A DNA hybridized to hexon mRNA (adapted from refs. [2] and [7]). Introns A, B and C of the DNA form loops because they are spliced in the mature mRNA. (*right*) Intron-exon organization of the Ad2 hexon gene. (*continued on next page*)

Fig. 1.1 | (*continued*) **b** (*left*) Pyrimidine and purine nucleosides are produced from cyanoacetylene (**1**)[8–10]. Cyanoacetylene reacts with hydroxyurea (**2**) and a sugar to give a ribosylated form of isoxazoleurea (**5**), a precursor of pyrimidines. In the presence of ammonia and nitrites, **1** reacts to hydroxyimino-malononitrile which combines with an amidine (**4**) and ribose upon reduction to ribosylated formamidopyrimidine (FaPy, **1**), a purine precursor. (*right*) Canonical Watson-Crick (WC) base-pairing that allows for templated self-replication. **c** Branch pathway of the two-step transesterification reaction common to group II introns and the spliceosome.[11–14] Displayed is the forward splicing reaction. The reversibility of both steps is exploited by mobile group II introns in retrohoming/retrotransposition. First, a bulged adenine in domain VI or U2 small nuclear RNA (snRNA) attacks the phosphate at the 5′-splice site. The cleaved 5′-exon, which still remains base-paired to EBS1/2, is then ligated to the 3′-exon whereupon the intron is liberated as a lariat. In either step, the incoming nucleophile may also be a water molecule (hydrolytic pathway) resulting in a linear intron and/or unligated exons. **d** Transition-state of the two-metal ion dependent phosphodiester bond cleavage by group II introns in the first step of splicing. Two divalent Mg^{2+} ions activate the incoming nucleophile (branched adenosine or water molecule) and stabilize the leaving group during the S_N2 type reaction.[15,16]

such as cyanoacetylene, hydroxyurea and ammonia, all considered to be readily available in the atmosphere and the crust of early Earth (Fig. 1.1b).[8,23] The necessary energy would come from electrical discharge while fluctuations in temperature (seasonal and day/night cycles) and pH (volcanic activity) and would allow for precipitation and enrichment of water-insoluble intermediates. All in all, it is not unlikely that prior to the "great oxidation" and the emergence of proteins, RNA had sophisticated catalytic and redox functionality.[24]

The unique combination of storage, self-copying and catalytic functionality seems to make RNA the ideal candidate for the development of life on the planet. Yet, an elaborately folded, three-dimensional structure, while being essential for chemical activity, inherently limits the speed at which the biomolecule is able to replicate. Over the course of evolution, the DNA double helix proved to be better for long-term storage of the increasing load of genetic information. Meanwhile, specialized proteins have taken over the functional aspects and are now the workhorses of the cell. Still, RNA remains not a molecular fossil of these past times, but continues to function as a regulatory hub between DNA and proteins.

As such, RNA orchestrates key processes including translation, splicing and gene expression.[25] It is believed that ribozymes like self-splicing group I and II introns are remnants of the RNA world. There is accumulating evidence that in eukaryotes, group II introns have evolved into a large protein-powered machine, the spliceosome.[26–28] This ribonucleoprotein (RNP) retains at its catalytic core a set of small nuclear RNAs (U2, U5 and U6 snRNA) which show considerable structural similarity to its self-splicing ancestor.[26] In the spliceosome, the 5′-exon is brought into the active site by base-pairing with the U5 snRNA. This long-range tertiary contact is mediated by the conserved exon-intron binding site 1 (EBS1-IBS1) interaction in group II introns. The kinetics of this particular interaction are examined in detail in chapter 4. Further analogies include the spliceosomal U6 snRNA which takes over the function of the catalytic domain V in group II introns, whereas the U2 snRNA contains the bulged adenine much like domain VI in the ancestral ribozyme (Fig. 1.2). Moreover, the spliceosome and group II introns share a common two-step transesterification mechanism which is assisted by two catalytic Mg^{2+} ions coordinated by residues in domain V or U6 snRNA (Fig. 1.1c and Fig. 1.2). Structural comparison of the accessory protein Prp8 in the spliceosome with intron-encoded maturases reveals extensive homology between their reverse-transcriptase and thumb/X domains.[26,29]

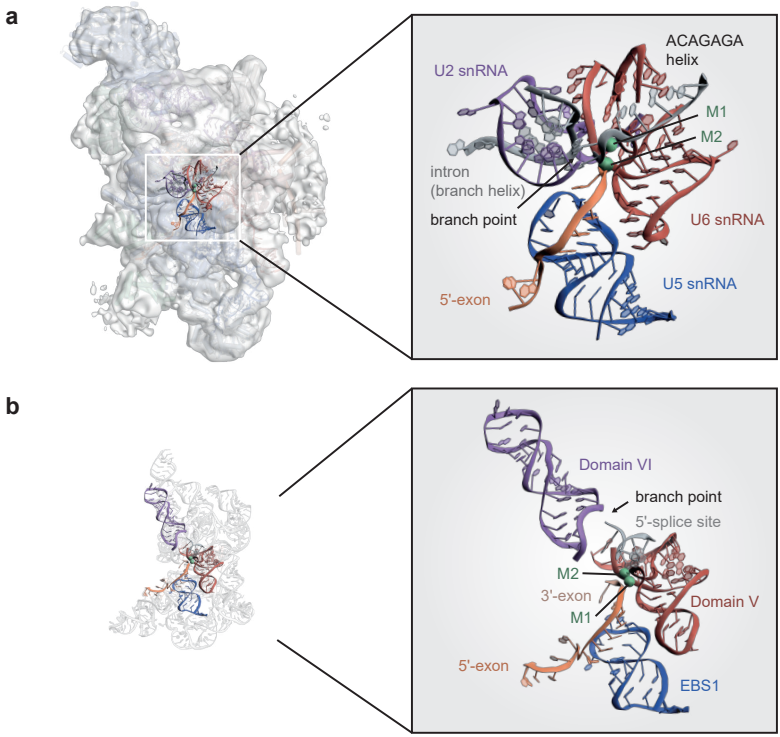


Fig. 1.2 | **Structural similarities between group II introns and the eukaryotic spliceosome suggest an evolutionary relationship.**
a Cryo-EM structure of the yeast spliceosomal C-complex and its catalytic RNA core (PDB: 5LJ5).[30] **b** Crystal structure of the branched group II intron lariat *LSUI2* from *Pylaiella littoralis*. Spliceosome and intron are drawn to scale to emphasize their difference in size. Corresponding RNA elements are highlighted in color. In the spliceosome the intron is secured by base-pairing with U2 (branch helix) and U6 (ACAGAGA helix) while the 5′-exon is held in place by the U5 snRNA. This interaction is recapitulated by the EBS1-IBS1 interaction in the group II intron. Two divalent metal ions (M1/2) stabilize the core and actively participate in the two-step transesterification mechanism (see also Fig. 1.1b/c).

The common domain architecture and similar interaction with RNA elements further suggest an evolutionary relationship. While the spliceosomal RNAs are cradled by Prp8, an extensive tertiary interaction network between domain I and the catalytic core provides stability in the group II intron. It is thought that this scaffolding domain I was gradually replaced by accessory proteins as the spliceosomal machinery evolved.[26]

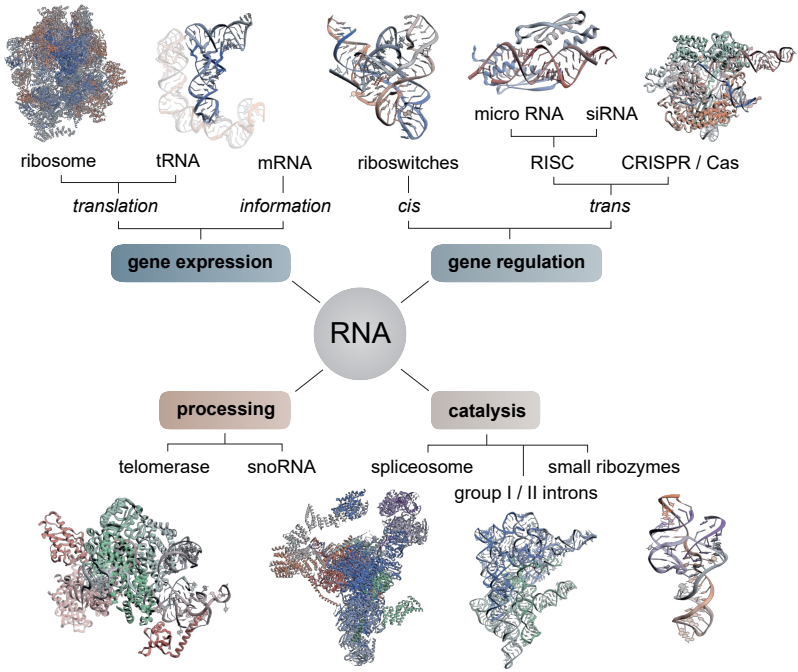


Fig. 1.3 | **RNA orchestrates a broad spectrum of cellular processes.** The multifaceted roles of non-coding RNA range from intron maturation and protein synthesis to mRNA degeneration and immunity (adapted from ref. [31]). The central dogma enunciated by Crick[17] concentrates on the gene expression branch. The work in this thesis focuses on catalytic group II introns and gene regulatory riboswitches.

1.3 RNA switches gene expression

Apart from self-splicing ribozymes, RNA are critically involved in numerous other cellular processes (Fig.1.3). Micro ribonucleic acid (miRNA) and small interfering RNA (siRNA) silence mRNA targets by forming a duplex which is then cleaved by a multiprotein assembly called the RNA-induced silencing complex (RISC).[32] RNA interference is a prime example where proteins and RNA work synergistically to fine-tune the levels of expression. However, there are also genetic control mechanisms that operate independently of protein cofactors. Here, the RNA acts as a molecular switch by sensing small-molecules such as thiamin pyrophosphate (TPP)[33,34], S-adenosyl methionine (SAM)[35,36] or flavin mononucleotide (FMN)[33,37]. Binding of the metabolite induces an allosteric structural change which activates or inactivates the downstream gene. These RNA sensors are known as "riboswitches".[38] Their "aptamers" (the metabolite binding domains), are often highly selective towards one particular signaling molecule, amino acid or nucleotide derivative (Fig. 1.4). Over the last twenty years many riboswitch aptamers have been identified, characterized and crystallized (Fig. 1.5).[39,40] Other mRNA leader sequences have long resisted biochemical validation.[41] Recently, a number of those riboswitches could be relieved of their orphan status and matched with their natural ligands.[40] Examples include the *ykkC* motif, which binds to the alarmone guanosine tetraphosphate (ppGpp) and the purine precursor phosphoribosyl pyrophosphate (PRPP) (Fig. 1.5c)[42–44], or the *yybP-ykoY* motif which was discovered as a potential riboswitch candidate early on[45], but only validated as a Mn^{2+} sensor[46,47] a few years ago. Although the bioinformatical identification of new consensus motifs gets ever more challenging due to their increasing rarity[48], there are nonetheless estimates of potentially hundreds if not thousands of yet unknown riboswitch classes spread over the bacterial domain.[39,41,49]

1.4 Vitamins bind RNA

When the first mRNAs capable of sensing the coenzymes B_1 (TPP), B_2 (FMN) and B_{12} (adenosylcobalamin, AdoCbl) were discovered in 2002[33,38,54,55], a new field of research was kicked off. Now, we look back on two "decades of riboswitches"[56]. One of the most astounding facts about these conserved RNA elements at the time was their abil-

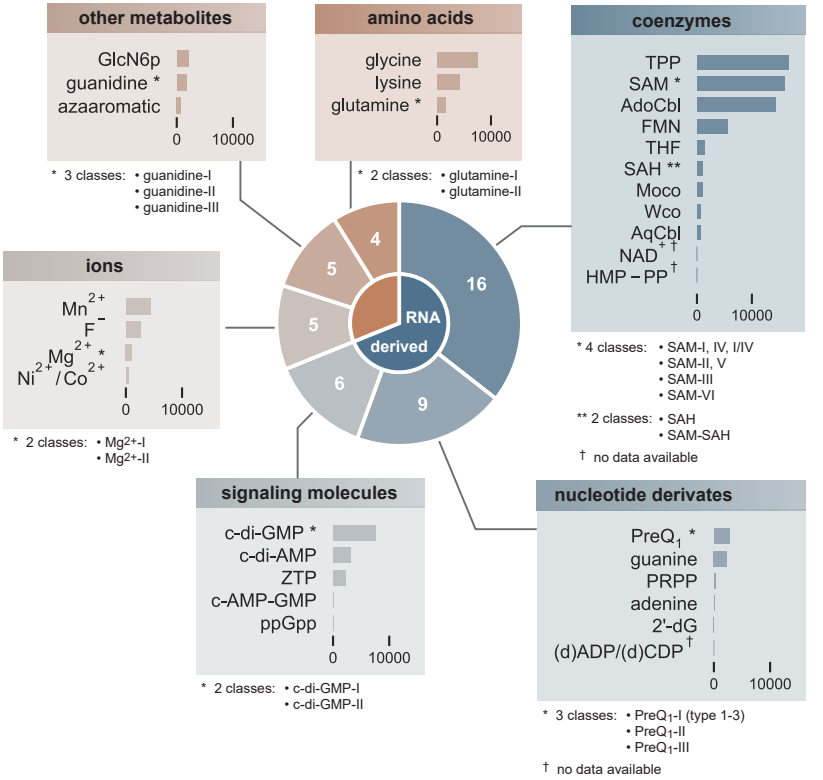


Fig. 1.4 | **Riboswitch distribution and their metabolites.**
Non-exhaustive list of known riboswitch classes named after their natural ligands.[39,40] The pie chart indicates the number of distinct riboswitch classes for each metabolite group. The bar plots show the relative abundance of the riboswitch class. TPP: thiamin pyrophosphate, SAM: *S*-adenosylmethionine, AdoCbl: adenosylcobalamin, FMN: flavin mononucleotide, THF: tetrahydrofolate, SAH: *S*-adenosylhomocysteine, Moco: molybdenum cofactor, WCo: tungsten cofactor, AqCbl: aquacobalamin, NAD^+: nicotinamide adenine dinucleotide, HMP-PP: hydroxymethyl-pyrimidine pyrophosphate, $PreQ_1$: prequeuosine$_1$, PRPP: phosphoribosyl pyrophosphate, 2′-dG: 2′-deoxyguanosine, c-di-GMP: 3′,5′-cyclic di-guanosine monophosphate, c-di-AMP3′,5′-cyclic di-adenosine monophosphate, ZTP: aminoimidazole carboxamide ribonucleotide, ppGpp: guanosine tetraphosphate, GlcN6P: glucosamine-6-phosphate. Not in the list is the T-box riboswitch binding to non-aminoacylated tRNA

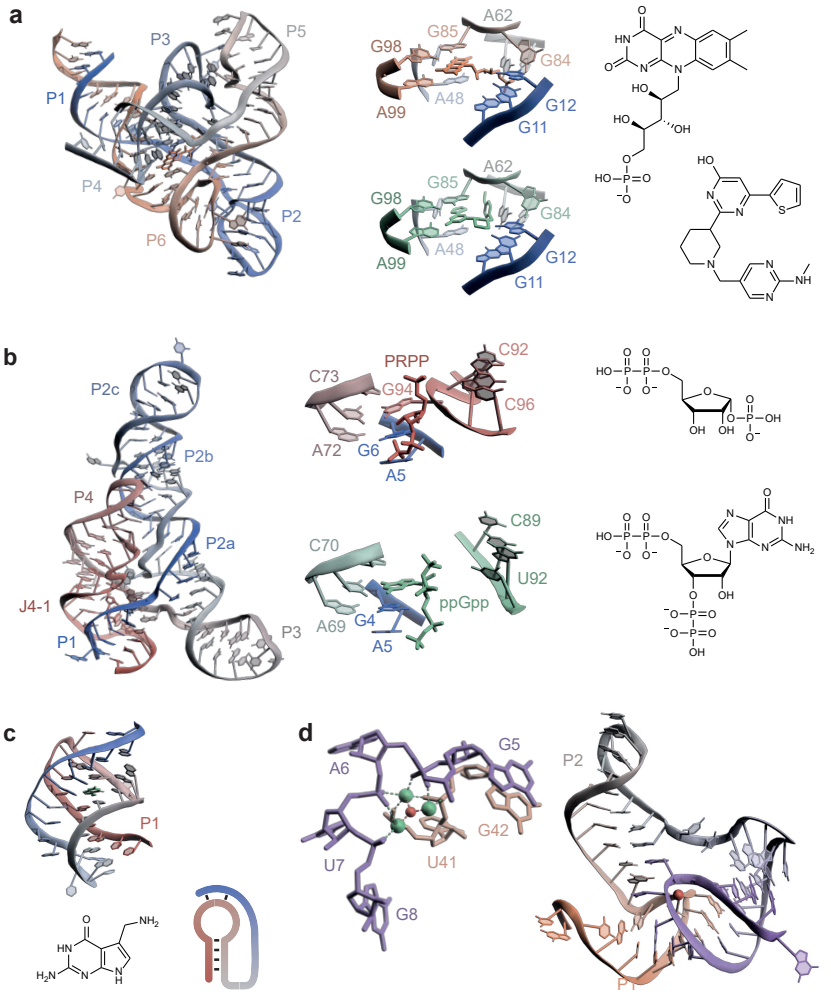


Fig. 1.5 | **Structural diversity of riboswitches aptamers.**
a FMN riboswitch from *F. nucleatum* (PDB: 3F2Q) in complex with flavin mononucleotide (orange, top, PDB: 3F2Q[37]) or ribocil (green, bottom, PDB: 5C45[50]). It is one of the few riboswitches which have been selected for the development of new antibiotics.[51] Ribocil inhibits riboflavin synthesis with an IC_{50} of 0.3 µM (*continued on next page*).[50]

Fig. 1.5 | (*continued*) **b** Structure of the *ykkC* riboswitch family bound to phosphoribosylpyrophosphat (PRPP, red, PDB: 6DLT[44]) or guanosine tetraphosphate (ppGpp, green, PDB: 6DMC[44]). Ligand specificity is achieved by a bulged guanosine in the PRPP riboswitch which forms a long-range interaction with C73. The ppGpp riboswitch misses the looped guanosine and instead uses the guanosine base of the ligand to contact C70. **c** Structure of the $PreQ_1$-I riboswitch (PDB: 3Q50[52]). Its relatively simple architecture consists of a H-type pseudoknot which embeds pre-queuosine$_1$, an intermediate in the biosynthesis of the nucleoside queuosine. **d** Architecture and metal binding core of the fluoride riboswitch (PDB: 4ENC[53]), the only aptamer discovered to date that selectively responds to a metal anion. The fluoride ion is encapsulated in a triangular cage of Mg^{2+} ions which is held in place by surrounding phosphate groups.

ity to recognize organic molecules without the assistance of proteins. Riboswitches regulate gene expression on the levels of transcription or translation.[49,57] The former involves rho-independent transcription termination through an intrinsic stem loop which is followed by a stretch of about eight uridines (Fig. 1.6a). These two sequence motifs stall the polymerase whereupon the inherently weak uracil/deoxyadenine-rich hybrid dissociates and the transcript is released from the complex.[58,59]
The second mode of action targets translation initiation whereby a sequestor hairpin prevents the ribosome from binding to the Shine-Dalgarno sequence located upstream of the start codon (Fig. 1.6b).[60–62] For B_{12} riboswitches both mechanisms have been proposed.[38,60–66] Gram-negative proteobacteria like *Escherichia coli* or *Salmonella typhimurium* have been found to preferentially repress gene expression by inhibiting translation initiation, whereas the gram-positive *Bacillus* and *Clostridium* species predominantly attenuate transcription.[65] In both cases, the mechanisms involve mutually exclusive RNA secondary structures in the so-called "expression platform" of the riboswitch downstream of the aptamer. These are the terminator/anti-terminator stem loops on the transcriptional level and the ribosomal binding site (RBS) sequestor / anti-sequestor hairpins in translation regulation.[49,57] The relative stability of these elements determines the genetic decision. In the cell, RNA starts to fold while it is still being actively transcribed. As a consequence, there is a temporal coupling between ligand binding and formation of regulatory elements. Depending on the availability and

affinity of the metabolite, the rate of aptamer folding, the length of the expression platform, and the speed of transcription, the riboswitch operates either under kinetic or thermodynamic control.[67] The switch is kinetically driven if the RNA polymerase (RNAP) has reached the genetic decision point *before* the metabolite and the aptamer are in equilibrium. In other words, the time for transcription of the expression platform is shorter than the time for metabolite equilibration: $\tau_{\mathrm{txn}} < \tau_{\mathrm{eq}} \sim 1/k_{\mathrm{off}}$. In this pre-equilibrium and co-transcriptional setting, the riboswitch is triggered only if the metabolite levels are significantly above the K_{d} determined in a post-transcriptional assay for the aptamer alone.[67] This is the case with the FMN riboswitch.[68] Because the dissociation time of FMN is slow with respect to the time needed for expression platform synthesis, the riboswitch acts much like a "molecular fuse".[69] Once the RNA commits to a particular fold it reaches a point of no return. Reverting the decision by transitioning from one kinetically trapped conformer to another would be simply too slow and the polymerase would have already run off.

Conversely, if the metabolite and the aptamer have enough time to equilibrate ($\tau_{\mathrm{txn}} > \tau_{\mathrm{eq}}$), the relative population of metastable and interconverting RNA structures ultimately determines the fate of the riboswitch-controlled proteins. The stability of alternate secondary structures is in this case thermodynamically controlled and may be modulated post-transcriptionally by metabolite binding.[67,73] Most riboswitches which regulate access to the ribosomal binding site are thus at least to some degree under thermodynamic control.[71,74,75] The *E. coli btuB* riboswitch, however, also shows that during transcription and folding of the expression platform the polymerase pauses at strategic sites in order for the much more complex aptamer to catch up.[76] Otherwise, competing sequences from the fast-folding expression platform (i.e. parts of the anti-sequestor) may interfere with the yet incompletely structured parts of the aptamer, thus preventing the possibility for the RBS-sequestor to form at all. From this perspective, the RNAP is buying time so that the aptamer fold can be consolidated by interacting with coenzyme B_{12}. In the absence of B_{12}, the anti-RBS strand invades the weak aptamer, once elongation resumes, and base-pairs with a complementary sequence to form the anti-sequestor.[76] In this way, RNAP pausing coordinates the co-transcriptional folding of the intricate aptamer with the comparatively simple architecture of the expression platform in a kinetically controlled scenario.

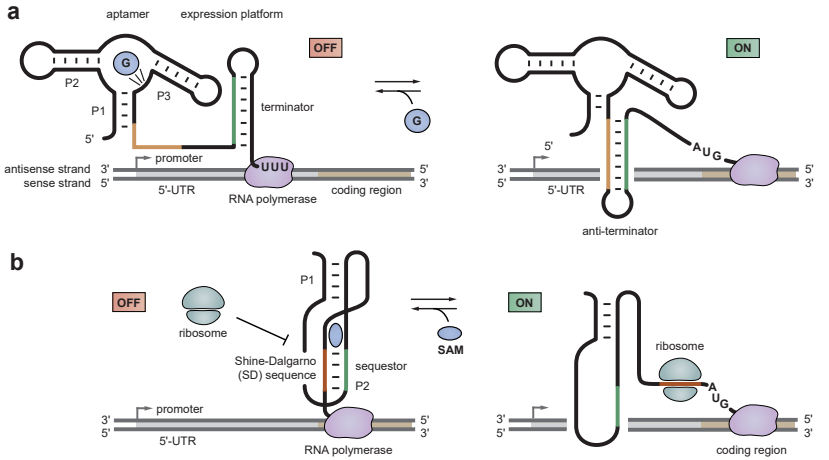


Fig. 1.6 | **Mechanisms of gene regulation by riboswitches.**
a The *xpt* guanine riboswitch of *B. subtilis* modulates premature transcription termination by formation of a intrinsic terminator stem. The RNA polymerase is stalled at a poly-U stretch and releases the nascent transcript. In the absence of guanine, part of P1 instead pairs with a complementary segment in the expression platform forming an alternative anti-terminator. The polymerase reads through and transcribes the full gene.[70] Interestingly, in the related *ydhL* adenine riboswitch of the same organism which shares a very similar aptamer topology, the directionality of regulation is reversed: adenine binding destabilizes the terminator stem in favor of the anti-terminator which is fully integrated into P1.[57,70] Instead of a repressor, the adenine aptamer thus acts as an on-switch. **b** SAM binding strengthens a pseudoknot in the SAM-II riboswitch whereupon the Shine-Dalgarno sequence (the recognition site for the ribosome) is sequestered, thus preventing translation initiation. When SAM levels are decreased, the pseudoknot is untied, exposing the ribosomal binding site and allowing the gene to be translated into protein.[71,72].

It is important to note that the concept of kinetic and thermodynamic control are simplified models and as such are not black and white. Both transcriptional and translational riboswitches can be thermodynamically driven[47,74,77–80], although quite a number of RNA that use terminator stems appear to be controlled by a resting RNAP and transcription factors such as NusA.[67,68,81–83]

1.5 Anatomy of a B_{12} sensor

The *btuB* genes in *E. coli* and *S. typhimurium* are critically important for cobalamin homeostasis.[60,61,64] They code for transporters that shuttle cobalamins across the outer membrane into the periplasm.[84] *Salmonella typhimurium* is additionally able to synthesize coenzyme B_{12} *de novo* using a number of proteins encoded by the *cob* operon.[85] Coenzyme B_{12} downregulates its own biosynthesis and availability as part of a negative feedback loop by sequestering the ribosomal binding site in the 5′-untranslated region (UTR) of the *btuB* or *cob* genes. With over 200 nucleotides in length, B_{12} responsive RNAs are amongst the largest and most abundant riboswitches identified to date (Fig. 1.4 and Fig. 1.7a).

Riboswitches have evolved to perform two main tasks, metabolite recognition and conformational switching. The natural design of B_{12} sensors reflects this dual functionality in the conserved aptamer spanning domains P1-P12 and the variable expression platform. The point of contact is a pseudoknot involving P5 and P13.[66] The aptamer is made up of two parallel coaxial stacks, the first formed by helices P1/P3/P6 and the second by regions P4/P5/P13 (Fig. 1.7c).[86,87] The stacks are connected by a T-loop (L4) docking into an internal loop between P6 and P7 (Fig. 1.7d).

In the related aquacobalamin riboswitch from a marine environmental metagenome (*env8* AqCbl, Rfam entry: RF01689[88]), this internal loop is replaced by a hairpin (L6) thus forming a T-loop/T-loop structure, a recurrent motif in riboswitches and ribosomal RNA.[37,86,89]. Additional long-range interactions between two adenines in J10/11 and the junction J6/3 pull the peripheral elements P7-12 back into the core (see also Fig. 1.10). This bent stem structure describes a U-shaped turn, as seen in the lysine riboswitch.[90] AdoCbl itself is embedded in a four-way junction with P4/P5 and P3/P6 framing the binding pocket (Fig. 1.7e). The corrin ring of the metabolite packs against the purine stack of J3/4 and

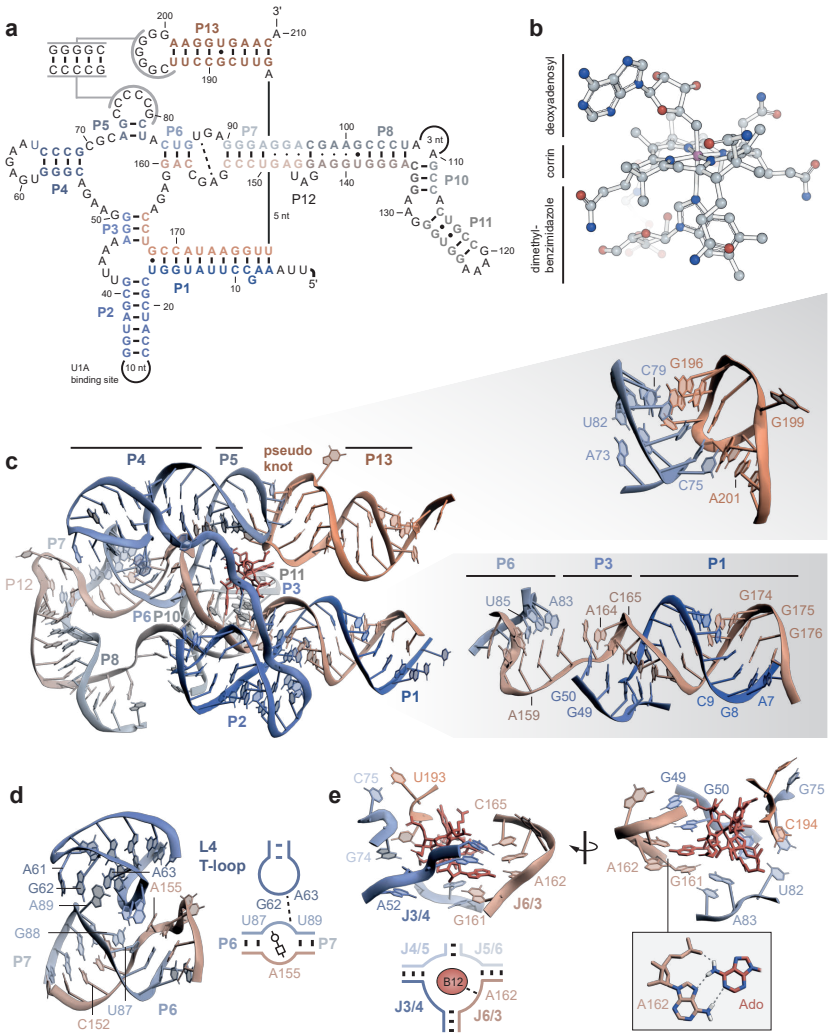


Fig. 1.7 | **Architecture of the adenosylcobalamin riboswitch.**
a Secondary structure of the *Thermoanerobacter tengcongensis (Tte)* AdoCbl riboswitch with base-paired regions P1-P13 highlighted in color. **b** Coenzyme B_{12} (adenosylcobalamin) coordinates a central Co^{+} ion by a corrin ring and two axial ligands, an α-dimethylbenzimidazole an the eponymous β-adenosyl group. Coenzyme B_{12} is the biologically active species of vitamin B_{12} and is formed upon exchange of a labile axial cyanide ligand. (*continued on next page*)

Fig. 1.7 | (*continued*) **c** Crystal structure of the *T. tengcongensis* AdoCbl riboswitch (PDB: 4GMA[86]) showing the two coaxial stacks P1-P3-P6 and P4-P5-P13 including a pseudoknot. The stacks are joined by a T-loop (L4) nested into an internal loop between P6 and P7. The latter partially takes over the role of a second T-loop as seen in the related aquacobalamin (AqCbl) riboswitch or the FMN riboswitch (see Fig. 1.5a). **e** The aptamer core consists of a four-way junction formed by P3-P6. Surprisingly, ligand recognition relies almost entirely on van der Waals interactions of the corrin side chains with nucleotides lining the binding pocket.[86,87] The metabolite is sandwiched between a stack of purines in J3/4 and J6/3, facing the β-axial ligand, and the pseudoknot on the opposite side. Only two hydrogen bonds are observed involving adenosine A162 in J6/3 and the adenosyl ligand of the cobalamin.

J6/3. The β-axial adenosyl ligand engages directly in a stacking and base-pair contact, explaining why the affinity of the *T. tengcongensis*, *S. typhimurium* and *E. coli* riboswitches is higher for AdoCbl than for methyl- and cyanocobalamins.[38,91]. In the *env8* AqCbl riboswitch, by contrast, an adenine is occupying the position of the adenosyl moiety, making this variant selective for cobalamins with sterically less demanding axial groups.[86] Cobalamins are the largest RNA coenzymes known to date. Binding to the RNA, which surprisingly relies mostly on van der Waals interactions, is thought to solidify the aptamer.[69,86] Although the importance of the kissing loop interaction has been postulated, the mechanism by which the binding signal is relayed to the expression platform is not fully understood and we try to address this in Chapter 5.

1.6 Ionic landscape architects

RNA and metal ions have been partners all along. Due to its polyanionic nature, RNA would be reluctant to fold upon itself without a cloud of counterions mitigating the electrostatic repulsion of the backbone phosphates.[92–95]. Without mono- and divalent ions interacting both diffusely and site-specifically with the RNA, the nucleic acid would be incapable of molecular recognition. Yet, the finding that a single or a few ions can also act as a "metabolite" and trigger an RNA rearrangement that translates into a genetic response was still somewhat unexpected.[96,97] The Mg^{2+} and F^- aptamers were only the first in a

row of riboswitches that sense atomic ions. Those include Mn^{2+}[46,80], Ni^{2+} and Co^{2+}[98] or F^-[53,99] responsive RNAs. The F^- aptamer displays a particularly unique coordination mode. The negatively charged fluoride ion is anchored by a cluster of three Mg^{2+} ions which in turn are surrounded by an outer shell of five non-bridging phosphate oxygens (Fig. 1.5d).[53] The structure demonstrates beautifully how ions are able to pack together distal parts of the RNA backbone and orient the bases into a pseudoknot-like architecture. Ion-responsive riboswitches are usually involved in metal ion homeostasis by regulating genes that code for the corresponding ion channels or transporters.[39] Binding cooperativity allows the RNA to respond to slight changes in the concentration level of these ions. This is particularly important for heavy metals which are toxic to the cell even at moderate concentrations. Their role as riboswitch cofactors highlights the intricate relationship between metal ions and RNA beyond charge compensation.

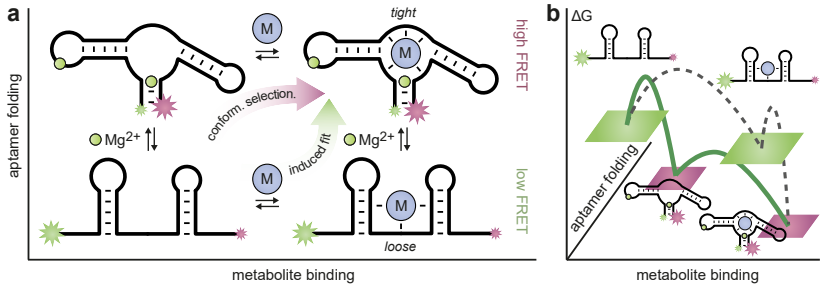


Fig. 1.8 | **Induced fit and conformation selection models for riboswitch dynamics. a** Four-state thermodynamic cycle illustrating two different pathways for a toy riboswitch interacting with a metabolite (denoted as M). Both the partially unfolded and the folded aptamer can recognize the metabolite. The conformational selection route posits that the ligand solidifies an already pre-structured aptamer. In the induced fit mechanism only binding of the metabolite provokes folding and tertiary contact formation. Single-molecule approaches can disentangle the different pathways by monitoring a possible formation of the folded aptamer upon addition of Mg^{2+} in the absence of metabolite. **b** The flux through the state network is dependent on the height of the energy barriers. The major route (conformational selection) for this hypothetical riboswitch is indicated in green.

Mg^{2+} riboswitches are special in a different respect. The elaborate tertiary structure of Mg^{2+}-I aptamers harboring multiple conserved Mg^{2+} binding sites allows the RNA to elicit a highly cooperative, "all-or-nothing" response.[39,97] While we are used to think of riboswitches as generic two-state switches, more recent single-molecule experiments and gene expression analyses suggest that this binary behavior of Mg^{2+}-I riboswitches might not necessarily be the norm. Instead, a number of RNAs, among those the abundant SAM-I, hydroxycobalamin and guanidine riboswitches, have been found to adopt a more fine-grained ensemble of conformations.[74,79,80,100] Unlike the traditional postulate of a simple two-state repression mechanism (apo: expression on, ligand bound: expression off), multiple conformers capable of binding the metabolite are likely to coexist in solution. In the simplest case, this is illustrated by a four-state thermodynamic cycle (Fig. 1.8a). The flux through this cycle is determined by the height of the energy barriers separating the individual states. Thus, the relative kinetic rates inform about whether ligand binding to the unwound conformation triggers folding (induced fit mechanism, "bind-then-fold") or alternatively, whether the ligand preferentially selects an already pre-folded structure from the ensemble and stabilizes it (conformational selection, "fold-then-bind").[73,101] Both pathways have been previously described for different riboswitches and rerouting has also been observed when changing the folding conditions.[71,74,80,101–103] Tuned by the abundance of natural ligand and metal ions, most riboswitches might thus act more like analog dimmers than digital switches.[104] The existence of many substructures is well in line with the general perception of a rugged energy landscape.[105–107] The roughness originates from the stability of individual base-pairs, particularly the canonical WC interactions.[108] This leads to many stable, alternate secondary structures each representing a local minimum on a furrowed free energy surface (Fig. 1.9). Metal ions, ligands and other environmental cues constantly remodel the shape of the energy basins and thereby redistribute the conformational ensemble.[107,109]
Divalent ions such as Mg^{2+} are particularly apt to bias dynamic RNA ensembles by selectively stabilizing certain motifs.[110] Mg^{2+} is well known to coordinate in an octahedral geometry to the phosphate groups of the RNA backbone and base nitrogens N7 (A/G) and N3 (C) either directly (inner-sphere) or mediated via a water molecule (outer-sphere).[111,112]. A number of metal ion binding sites have been located at the interface of independently folded segments of RNA secondary structures in the

ribosome[113] as well as group I and II introns.[114,115] Much like the ion cluster in the small fluoride riboswitch, these structural ions help to assemble and stabilize contacts between individual domains. The kinetics of such a long-range contact are explored in Chapter 4. Recently, amino acid chelated Mg^{2+} has also been shown to enhance folding and catalysis.[116] Similar to diffuse, divalent ions, those weakly bound Mg^{2+} are driving RNA compaction. Metabolome analyses indicate that RNA in the cytosol is exposed to more than hundred millimolar of amino acids, in particular glutamate.[117] Along with nucleotide triphosphates such as ATP, these metabolites scavenge most Mg^{2+} in the cell, leaving only a small fraction of ions (about 0.5 mM–1 mM in prokaryotes and 1.5 mM–3 mM in eukaryotes) free and fully hydrated.[1] The comparatively high abundance of weakly chelated Mg^{2+} might thus compensate for the low physiological levels of free Mg^{2+} and contribute to compact RNA intermediates to a previously underestimated degree.[116]

1.7 Parallel routes to function

How does an RNA molecule navigate the folding routes to its functional form? The challenge for RNA is two-fold[119]: The first is one of *kinetics*. RNA molecules tend to get trapped in misfolded conformers due to the many favorable interactions between nucleotides at their Watson-Crick, Hoogsteen or sugar edges.[120] This may even be exacerbated by the presence of metal ions. Incorrectly folded secondary and tertiary structures drop out of the pool of functional molecules unless they are rescued for instance by a RNA chaperone.[121] The second, related issue is of *thermodynamic* nature. As there are many secondary structures similar in shape and free energy, the RNA may struggle to discriminate between incorrectly assembled helical bundles and wrongly docked loops. This opens up many parallel routes and introduces heterogeneity in the folding kinetics.[108,122,123] For RNA the existence of such ensembles can also turn into an advantage since dynamic sampling of multiple conformations is integral to biomolecular function[109]. But it is a serious complication for modeling RNA structure (see Section 1.8: Can we predict RNA structure from scratch?).

[1] For comparison, the the total Mg^{2+} concentration is around 20 mM–50 mM depending on the cell type.[116,118]

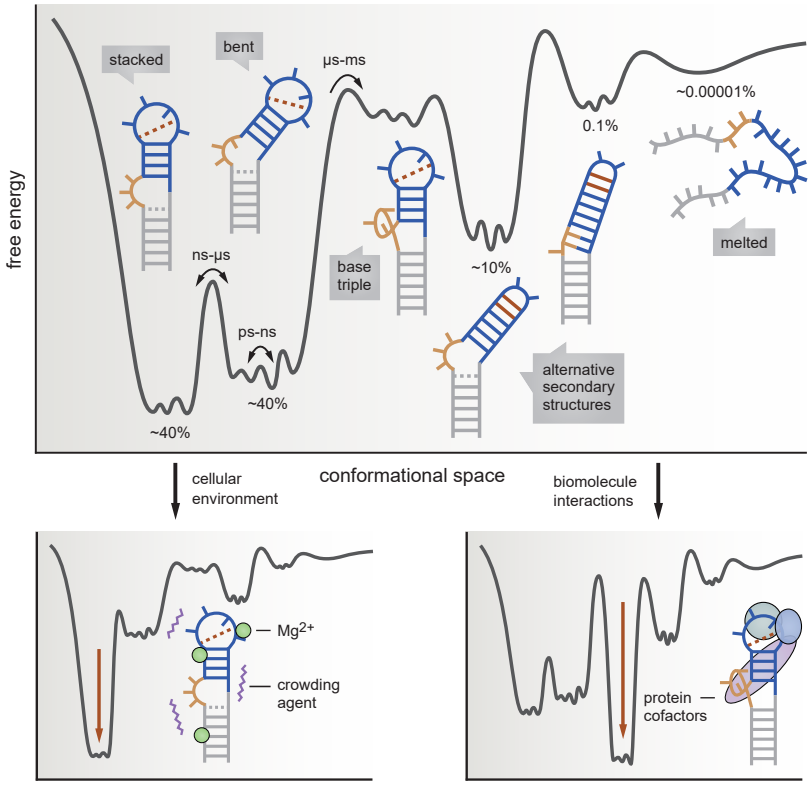


Fig. 1.9 | **Binding interactions modulate the free energy surface and redistribute the conformational ensemble.** Free energy profile of the trans-activation response element (TAR) of HIV-1 (adapted from ref. [109]). The local minima are populated by alternative secondary structures which interconvert at different rates depending on the height of the barrier separating them. Cellular components such as metal ions, crowding agents and other macromolecular cofactors remodel the energy landscape to favor certain conformations.

The unifying principle across the many types of RNA is that their folds are governed by base stacking and base pairing.[124] RNA structures are organized in recurrent modules (Fig. 1.10), which are interconnected by helices and pinned together by non-canonical base-pairs.[125] While there is a hierarchy to RNA structure dictated by energetics[126] – secondary structures with WC base pairs are the most stable elements and also tend to fold first – there is extensive cooperativity among individually pre-assembled motifs, especially at later stages of the folding process.[108,127] This means that formation of one tertiary contact favors formation of another. In the *Azoarcus* ribozyme these cooperative tertiary networks emerge surprisingly early, blurring to some extent the notion of folding hierarchy.[128] In the current view, RNA folding proceeds through collapsed intermediates whose formation is assisted by metal ions[108,123,129,130], molecular crowders[131,132] and RNA binding proteins[119,133–135]. The rate-limiting step is the search for stable tertiary contacts which may prove to be cooperative or antagonistic (negative cooperativity). In light of the structural modularity of RNA, efforts have been devoted to building a quantitative, predictive model for the stability of RNA tertiary structures, akin to those available for RNA secondary structure.[136,137] This so-called "reconstitution model" states that the folding free energy of a module ΔG_{fold} can be separated into independent, additive terms.[138,139] For two helices connected by a junction (HJH) and contacting each other via a tertiary interaction, the contributions include the energetic cost for aligning flanking helices ΔG_{align} (with a correction factor accounting for the particular ionic environment $\Delta G_{+/-}$) and the favorable energetics for forming the tertiary contact ΔG_{tert}.

$$\Delta G_{\text{fold}} = \Delta G_{\text{align}} + \Delta G_{\text{tert}} = \Delta G_{\text{HJH conf.}} + \Delta G_{+/\text{-}} + \Delta G_{\text{tert}} \quad (1.1)$$

Importantly, ΔG_{fold} is computed for module *ensembles*, not just a single crystal structure, and therefore gives rise to free energy distributions. The reconstitution model builds on single-molecule as well as high-throughput characterizations of such motif ensembles. Specifically, different helices and junctions are inserted in various structural contexts to generate a thermodynamic fingerprint for each motif.[140,141] The concept of "ensemble additivity" suggests that a convolution of ensembles determined from isolated modules can predict the folding stability of an arbitrary RNA. Cooperativity among motifs can be readily accounted for by iteratively recalculating the conformational ensemble and the as-

sociated energy term $\Delta G_{\text{HJH conf.}}$ as the number of interacting modules grows. Overall, the reconstitution model is a first step towards a quantitative prediction of RNA folding that goes beyond qualitative analogies such as the one of rugged energy landscape laden with kinetic traps.[139]

1.8 Can we predict RNA structure from scratch?

RNA is quite flexible. Its sugar backbone contains six dihedral angles compared to only two in proteins. The additional degrees of freedom increase the phase space of RNA, which is reflected in larger RMSDs of structural models. Relatively small torsions at the core can induce large hinge motions at the periphery, such that emanating stem loops can reorient quite substantially. The hierarchy of RNA folding posits that the secondary structure (helices) is stable without tertiary contacts. The energetics of helices containing standard WC base pairs are well established and described as a set of nearest neighbor parameters sometimes referred to as the "Turner rules".[136,137] They are mostly derived from optical melting experiments and are integrated in many of the most popular secondary structure prediction tools like RNAstructure[142], Mfold[143] or ViennaRNA[144]. With these algorithms and additional data from chemical probing or phylogenetic analyses it is relatively straightforward to annotate and model helical regions. The challenge is to predict where and how these helices come together to form the compact architectures seen in crystal structures. One such contact point are pseudoknots, a feature common in riboswitches and ribozymes. While dynamic programming algorithms generally scale poorly when trying to predict pseudoknots, combined with a heuristic they are making progress in locating these elements, which are key to the overall topology.[137] Pseudoknots are still largely based on WC interactions. The real problem comes with tertiary interactions that rely on non-WC base pairs such as A-minor interactions[145] or ribose zippers[146] (Fig. 1.10a). A series of blind challenges, called RNA-puzzles and inspired by the Critical Assessment of Methods for Protein Structure Prediction (CASP) trials for proteins[147], have not unexpectedly identified non-WC base pairs as the main bottleneck for accurate RNA structure prediction (Fig. 1.10b).[148–151] It is also these non-canonical interactions that make up many of the conserved RNA motifs including C-loops[152], G-bulges[153,154] or K-turns[155] (Fig. 1.10c). The prospect of predicting RNA structure from nothing but the primary sequence is still quite far away. Still, considerable progress has been

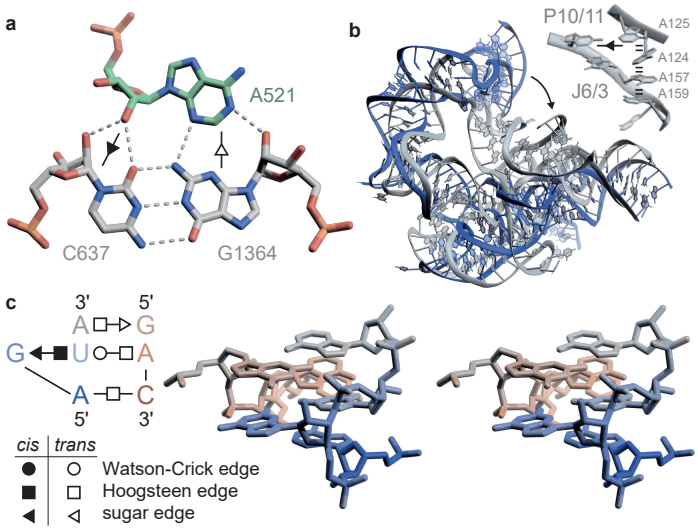


Fig. 1.10 | **RNA structure is modular.**
a Type I A-minor interaction in the 50S ribosomal subunit from *H. marismortui*.[145] A-minor motifs are the most widespread tertiary interactions in the ribosome and amongst the most important long-range contacts in large RNAs. **b** Fragment Assembly of RNA with Full-Atom Refinement (FARFAR2) prediction of the cobalamin riboswitch (RNA-puzzle 6[156], blue) against the crystal structure (PDB: 4GXY[87], gray). The RMSD of the top 1% models reaches 10 Å against the native structure. The cobalamin riboswitch is one of the hardest problems because of the extensive tertiary structure network in the aptamer and the non-inclusion of the metabolite in the modeling.[149] The model nicely illustrates the importance of non-WC interactions: most notably, the prediction misses the long range interaction between adenines A124/125 and the J6/3 junction, anchoring the peripheral domains P10/11 to the core (arrow). The interaction involves a *cis* sugar-sugar-edge contact and a continuous four adenine stack (dashed line). **c** The G-bulge module is a triple base-pair internal loop motif which occurs in the conserved sarcin-ricin loop (SRL)[157] in bacterial 23S rRNA and the eukaryotic loop E[158] of 5S rRNA. The module is annotated with the nomenclature by Leontis and Westhof[120] which organizes all pairwise interactions between any of the four nucleotides into 12 families each consisting of a 4×4 matrix. Cross-eye stereo representation of the SRL in *E. coli* 23S rRNA (PDB: 3DVZ).

made in the design of the algorithms, many of which are implemented nowadays as standalone and automated web servers.[159,160] They allow incorporation of experimental restraints from various sources such as chemical mapping experiments[161] or cryo-EM[162]. These developments have also been specifically encouraged in previous rounds of the RNA puzzle trials.[149,150] Puzzles where prior knowledge about local geometries, for example from structural homology, could be included in the modeling generally scored better.[150] Local templates help to assemble multiple helices, which remains challenging in purely *de novo* scenarios as evident from their looser packing (Fig. 1.10b).[149,156] Algorithmic advances including a stepwise Monte Carlo (SWM) method implemented in Rosetta and combined with improvements to the scoring functions have recently led to predictions of a variety of isolated non-canonical motifs at atomic resolution.[163] Similarly, better base pair sampling and filtering of poor conformations, together with an updated crystallographic library, has allowed Rosetta's fragment assembly protocol, FARFAR2, to recover native-like structures for many of the previous RNA puzzles.[156] Lack of sufficient conformational sampling is still a major issue for problems larger than 80 nt. The RNA puzzle trials have also shown that it is non-trivial to correctly rank the computed models[124,150]. This is commonly done through clustering.[164] Better refinement strategies to avoid sterical clashes as well as inclusion of small-molecules to improve packing and to model riboswitches dynamics also remain open challenges.[124]

Summary: RNA at the origin of life?

- The ability of RNA to self-replicate as well as the evolutionary trails of autocatalytic ribozymes and riboswitches argue in favor of the **RNA world hypothesis**. Insights from both structural biology and pre-biotic chemistry will continue to provide new perspectives into how life may have started on earth.
- **Riboswitches** alter gene expression through a positive or negative feedback loop. Oftentimes, the regulated protein is involved in membrane trafficking or biosynthesis of the small-molecule that binds to the riboswitch. A strikingly high number of metabolites are themselves precursors or derivatives of RNA building blocks. The *btuB* riboswitch of *E. coli* was amongst the first and largest *cis* regulatory RNA elements to be discovered and folds into a complex tertiary structure that selectively recognizes coenzyme B_{12}.
- The fate of nucleic acids is tightly linked to metal ions. They direct RNA folding by stabilizing individual motifs, bridging strands and securing labile non-canonical base pairs. Many ribozymes, including group II introns, are **obligate metalloenzymes** which accommodate one or several Mg^{2+} ions in their active core to enhance catalysis.
- Research on RNA folding has gone through several "epochs"[139]. It started with the realization that RNA is organized hierarchically and led to the systematic classification of many recurrent modules that are key to RNA's tertiary architecture. Efforts are now directed towards a **generalizable folding model** that allows to make quantitative predictions about the stability of three-dimensional RNA structures.

1.9 Research objectives

The work described in this thesis interfaces the fields of nucleic acid chemistry, biophysics and structural biology. It focuses on two main RNA molecules, a riboswitch and a ribozyme, which were investigated by single-molecule fluorescence spectroscopy. Different bioinformatical tools were built along the way to integrate FRET experiments and simulations.
In the following, we start with a short overview of the most important principles in fluorescence spectroscopy and the state-of the-field with a perspective on RNA dynamics and FRET. The subsequent chapters describe our published and unpublished work on RNA labeling, heterogeneity and modeling. Some of the key questions addressed in each chapter are outlined below:

Chapter 3: RNA labeling – putting a light on RNA

Development of a new labeling method for RNA faces three main challenges:

- How to **selectively** target a single base given the limited alphabet of four nucleotides with overall similar reactivity?
- How to overcome the **size** and **site constraints** of the current gold standard, that is chemical synthesis combined with enzymatic ligation?
- How to preserve **folding** and **activity** of the RNA when introducing two bulky fluorophores?

Chapter 4: RNA heterogeneity – exploring the kinetics of exon recognition

Group II introns engage long-range tertiary contacts for forward and reverse splicing. Here, we study the mechanisms of molecular recognition prior to the first transesterification reaction.

- What are the thermodynamic and kinetic differences between **RNA-DNA** and **RNA-RNA** contacts?

- What is the **structural origin** of kinetic heterogeneity and FRET state degeneracy?
- How do mono- and divalent metal ions **fine-tune exon binding and release** in group II introns?

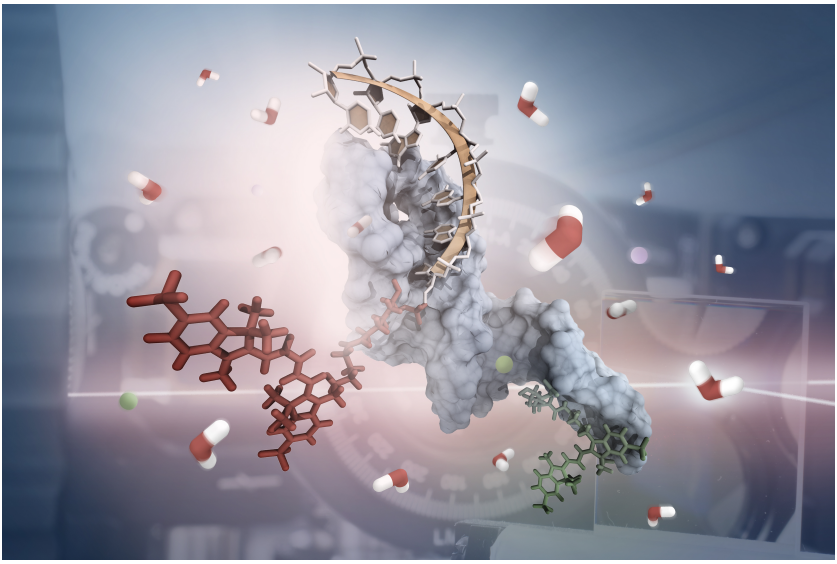
Chapter 5: RNA modeling – predicting the structure of a riboswitch

Riboswitches couple metabolite sensing to conformational rearrangements. How the binding signal is communicated to the expression platform often remains unclear though.

- How to best **predict the structure** of the full length *btuB* riboswitch of *E. coli*?
- Can single-molecule **FRET** experiments **refine** a modeled riboswitch structure?
- How is aptamer **folding coordinated** with the expression platform?

Chapter 2

Fundamentals of Biophysics



Molecular Dynamics

A single, fluorescently labeled RNA illuminated by a laser beam. Simulation and experiment cast a light on the wiggling of biomolecules.
– SNF Image competition (2017)

2.1 Absorption and Fluorescence

One of the most sensitive measurements in physical sciences is the detection of light. Against a dark enough background, nearly every single photon from a luminescent emitter can be registered. Molecules that release a photon after being excited by a light source are called fluorophores. The fluorescence quantum yield relates the number of emitted to absorbed photons

$$\phi_F = \frac{N_{\text{emitted}}}{N_{\text{absorbed}}} = k_F \int_0^\infty e^{-t/\tau} dt = k_F \tau_F = \frac{k_F}{k_F + k_{nr}} \tag{2.1}$$

To keep this ratio close to unity, bright fluorophores such as rhodamine 6G maximize the relative rate of fluorescence k_F with respect to other possible deactivation pathways. These include non-radiative decay routes such as internal conversion or intersystem crossings to spin-forbidden triplet states (Fig. 2.1a). Non-radiative processes compete with fluorescence and therefore decrease the brightness of the dye. Vibronic transitions are most likely to occur when there is little change in the nuclear coordinates. This is visually represented by two slightly shifted, asymmetric potential energy wells (Fig. 2.1b). In this semi-classical picture a transition is vertical if the wave functions of ground and excited state are in phase. The square of the overlap integral, the so-called Franck-Condon factor, is proportional to the transition probability. A typical excitation-emission cycle proceeds as follows: absorption of a photon induces a change to the electronic structure of the fluorophore whereupon an electron is promoted to a higher energy level. As the dye looses some of its energy again due to thermal dissipation, predominantly by collisions with solvent molecules, the electron drops down to the lowest vibrational level of the first excited state (Kasha's rule[165]). From there it will decay to the ground state while emitting a photon. The time the molecule stays in the singlet state S_1 is referred to as the fluorescence lifetime which can be measured by time-correlated single photon counting (TCSPC) (Fig. 2.1b). The fluorescence lifetime defines the time window during which dynamics of the molecule can be observed. It is related to the quantum yield through the rate of fluorescence k_F (Equation 2.1). A number of dyes are known to be sensitive to their physicochemical environment, such as viscosity, pH, temperature or solvent polarity.[166] Their lifetimes are susceptible to quenching or, as in the case of carbocyanines, fluorescence enhancement. The π-system

of this class of fluorophores extends across a series of isomerizable, conjugated double bonds.[167] It has been shown that stacking interactions with protein side-chains and nucleic acids residues reduce *cis-trans* isomerization which in return lowers the probability of non-radiative deactivation.[168–171] Protein and RNA induced fluorescence enhancement (PIFE/RIFE) can thus be exploited as a proximity sensor[170,172], complementing other short-range quenching methods like photoinduced electron transfer (PET).[173]

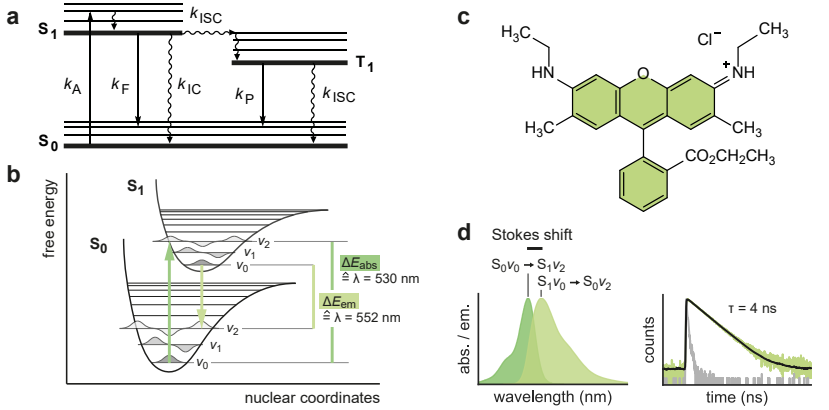


Fig. 2.1 | **Dye photophysics and its spectroscopic fingerprint.**
a Jablonski diagram displaying the main absorption and relaxation events (S: singlet state, T: triplet state, k_A: absorption, k_F: fluorescence, k_P: phosphorescence, k_{IC}: internal conversion, k_{ISC}: intersystem crossing. Panel adapted from ref. [173]). **b** Schematic potential energy diagram of the ground and first excited states. The coordinates of the two energy wells favor transitions between vibrational levels $v = 0$ and $v = 2$. The transition lines are inhomogeneously broadened do to fluctuations of the solvent shell around the fluorophore.[173] The small asymmetry of absorption and emission spectra is caused by slightly different relative probabilities of the vibronic transitions. **c** Chemical structure of rhodamine 6G. With its comparatively rigid dye scaffold, rhodamine 6G has an exceptionally high quantum yield of 0.95 (in ethanol[174]). **d** Absorption and emission spectrum of rhodamine 6G with a Stokes shift of 22 nm and a fluorescence lifetime of 4 ns.

2.2 Why single molecules?

The fluorescence signal measured in a conventional spectrometer originates from many copies of the same molecule. Assuming that this molecule exists in both folded and unfolded conformations and these differ with respect to some property (e.g. inter-dye distance measured via FRET, see below), the observed signal will be the weighted mean of the two subpopulations. In case of FRET, this is likely to result in an artificial distance that has no structural basis. Single-molecule experiments, in contrast, avoid this ensemble averaging by collecting photons from every molecule individually. In this way, subpopulations can be separated given that the transition rates are slow with respect to the time resolution of measurement. A biological system where multiple conformers or morphologies coexist is called heterogeneous.[175–177] Heterogeneity can be classified as either static (i.e. different structural states) or dynamic (i.e. different interconversion rates) and manifests itself in the ruggedness of the energy landscape.[107,178] While a biomolecule folds and functions many local minima are being populated. Ribozymes, for instance, often change conformation during their active cycle. To follow the time course in an ensemble experiment, all molecules would have to be synchronized. Such a trigger could be the addition of Mg^{2+}.[179] On the single-molecule level, changes over time can be readily obtained from equilibrium measurements without any perturbation. Dynamics are observed by monitoring fluorescence trajectories of immobilized molecules (Fig. 2.2a) or by correlating the intensity fluctuations of freely diffusing molecules (Fig. 2.2b). In either case, the challenge is to isolate the fluorescence signal of a single molecule in a sea of solvent molecules, which all contribute to background scattering. To reduce the amount of stray light, the observation volume has to be confined as much as possible. In practice, spatial selection is achieved by focusing a laser beam to a diffraction limited spot, the confocal volume, or by generating an evanescent field via total-internal reflection. A large Stokes shift of the fluorophore facilitates the spectral separation of the excitation and emission light and thereby improves the signal-to-noise ratio.
In camera-based total internal reflection fluorescence (TIRF) imaging, time traces of many molecules can be recorded in parallel.[180] Dynamics slower than about 10 ms can be extracted directly from the state dwell times.[178,181–183] Point detectors extend the observable dynamics to the sub-millisecond range.[184–186] The higher time resolution comes at the

price that molecules need to be interrogated sequentially and dynamics are usually limited to the translational diffusion time. Dynamic processes that are faster than the time it takes the molecule to diffuse through the confocal volume can be inferred from the time-correlated single photon statistics. A variety of methods have been developed to detect conformational exchange including burst variance analysis (BVA)[187] or photon distribution analysis (PDA).[188,189]

Fluorescence correlation spectroscopy (FCS) complements FRET by providing information on the photophysics (triplet state blinking, antibunching) as well as on translational and rotational diffusion. The timescales covered by FCS reach from picoseconds up to milliseconds or even seconds by making use of the recurrence effect[192]. In the nanosecond range FCS probes intramolecular chain reconfiguration dynamics which can be used to characterize the unfolded state ensemble of biomolecules.[186,193,194] The choice of instrumentation and analysis is thus ultimately dictated by the system under study. Proteins folding at the "speed limit"[195] naturally require fast electronics. Likewise, intrinsically disordered proteins usually exhibit dynamics in the submicrosecond range.[196] Folding of nucleic acids often proceeds through multiple intermediates separated by distinct energetic barriers and therefore typically occurs on the timescale of milliseconds to seconds.[108] Consequently, RNA folding and function is more amenable to camera-based detection which is reflected by a majority of studies using this detection scheme.[74,123,129,134,135,178,197]

2.3 A spectroscopic ruler at the nanoscale

In 1948, German physicist Theodor Förster postulated that the efficiency of energy transfer E from a donor to an acceptor fluorophore is proportional to the inverse sixth power of the inter-dye distance r.[198]

$$E = \underbrace{\frac{1}{1+\left(\frac{R_{\mathrm{DA}}}{R_0}\right)^6}}_{\text{(i)}} = \underbrace{\frac{I_{D,ex}^{A,em}}{I_{D,ex}^{A,em}+\gamma I_{D,ex}^{D,em}}}_{\text{(ii)}} = \underbrace{1-\frac{\tau_{\mathrm{DA}}}{\tau_{\mathrm{D}}}}_{\text{(iii)}}. \tag{2.2}$$

The distance where exactly 50% of the donor excitation energy is transferred to the acceptor is called the Förster radius, R_0 (Fig. 2.3). This quantity is experimentally derived from the donor quantum yield ϕ_D,

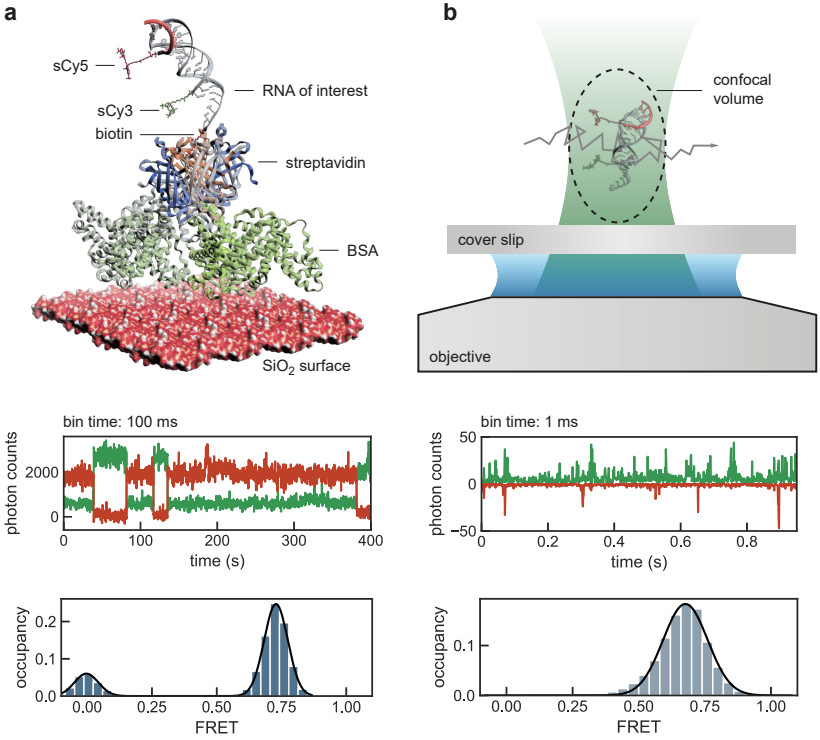


Fig. 2.2 | **Imaging single molecules by surface immobilization and free diffusion.**
a RNA hairpin labeled with a pair of fluorophores and immobilized via a streptavidin-biotin linker on a quartz (SiO_2) slide which is coated with bovine serum albumin (BSA). Single RNA molecules are imaged on an electron multiplying charge-coupled device (EMCCD) and show anti-correlated donor and acceptor signal resulting in a donor only FRET state and a bound state around 0.7. **d** Confocal illumination of freely diffusing molecules. While transiting though the Gaussian shaped confocal volume, the molecules emit a burst of photons that is spectrally separated and collected on a single-photon avalanche diode (SPAD). Donor-only molecules are filtered out by applying an intensity threshold of $F_{\mathrm{A}_{\mathrm{ex}}}^{\mathrm{A}_{\mathrm{em}}} > 20$ photons. The histograms are corrected for donor leakage into the acceptor channel, direct excitation of the acceptor, differences in quantum yields of donor and acceptor dye and detection efficiencies of the color channels, as described in detail in chapter 5 and in the literature.[190,191]

the spectral overlap of the donor emission F_{D} and the acceptor excitation spectrum ϵ_{A}, the relative orientation of the dye dipoles κ^2 as well as the refractive index n of the intervening medium

$$R_0 = \left(\frac{9000 \ln(10)\, \kappa^2\, \phi_D}{128 \pi^5\, N_A\, n^4} \int_0^\infty F_{\mathrm{D}}(\lambda) \epsilon_{\mathrm{A}}(\lambda) \lambda^4 d\lambda \right)^{1/6} \tag{2.3}$$

where N_A is Avogadro's constant.[166] Most of these parameters are dependent on the environment of the fluorophores and will therefore affect the accuracy of FRET as a spectroscopic ruler. (1) The donor quantum yield is sensitive to quenching and photoisomerization and thus often changes upon labeling. (2) Like the quantum yield, κ^2 is dependent on the microenvironment of the fluorophore. If the mobility of the dye is constrained, the value of κ^2 deviates from the isotropic average of 2/3, where dye reorientation is assumed to be fast with respect to the rate of energy transfer.[199] The uncertainty of κ^2 can be estimated by fluorescence anisotropy decays.[200] Trapping of fluorophores on the biomolecular surface, indicated experimentally by a residual anisotropy after depolarization, has also been observed in molecular dynamics simulations[171,201–203] and is included in the most recent, refined dye models.[171,204,205] (3) Absorption and emission spectra of bioconjugated fluorophores tend to shift towards longer wavelengths compared to the free dyes and finally (4) adapting the refractive index n to the local medium (protein, nucleic acid) embedding the donor and acceptor dyes may be more appropriate than using the value for the solvent.[206]

The FRET efficiency is calculated either by correlating the photon intensities of donor and acceptor dye (Equation 2.2 ii) or by comparing the donor lifetimes in the presence and absence of an acceptor (Equation 2.2 iii). In intensity-based measurements the factor γ corrects for the mismatch in quantum yields ϕ of the dyes and detection efficiencies η of the two channels: $\gamma = \phi_{\mathrm{A}} \eta_{\mathrm{A}} / \phi_{\mathrm{D}} \eta_{\mathrm{D}}$. Methods to quantify γ have been proposed for confocal as well as surface-immobilized experiments.[190,207]

Equation 2.2 i relates the FRET efficiency to a single fixed donor-acceptor distance R_{DA}. This is valid for a static structure and rapid sampling of fluorophore orientations with respect to the donor excited state lifetime (i.e. $\kappa^2 = 2/3$). Yet, biomolecules are dynamic and the dyes themselves are attached via flexible linkers. Thus, the mean FRET efficiency $\langle E \rangle$ is averaged over a distance probability distribution $P(R_{\mathrm{DA}})$

$$\langle E \rangle = \int_0^\infty P(R_{\mathrm{DA}}) \frac{1}{1 + \left(\frac{R_{\mathrm{DA}}}{R_0}\right)^6} dR_{\mathrm{DA}} \tag{2.4}$$

A number of models from polymer physics have been utilized to describe the distance distribution $P(R_{\mathrm{DA}})$ including a Gaussian chain[208–210], a worm-like chain[211] or a self-avoiding walk.[212–214] In the Gaussian chain model the distance distribution is characterized by a single parameter, the mean-squared end-to-end distance $\langle r^2 \rangle$

$$P(r) = 4\pi R_{\mathrm{DA}}^2 \Big(\frac{3}{2\pi \langle r^2 \rangle} \Big)^{3/2} e^{-\frac{3R_{\mathrm{DA}}^2}{2\langle r^2 \rangle}} \tag{2.5}$$

The fitting parameter $\langle r^2 \rangle$ is related to the radius of gyration $R_{\mathrm{g}} = \sqrt{\langle r^2 \rangle / 6}$ and can thus be compared to small-angle X-ray scattering (SAXS) experiments. Alternatively, distance distributions may be obtained from all-atom or coarse-grained molecular dynamics by simulating diffusion in a potential of mean force.[212–214] On top of such a trajectory, photon emission events can be sampled in a Monte-Carlo type simulation. By including the photophysics in this way, namely the transition rate matrix $\mathbf{K}_0(r)$ and steady-state populations $\mathbf{p}_{\mathrm{ss}}$, various properties of the photon statistics can be calculated. For example the donor correlation function is given by

$$g_{\mathrm{DD}}(\tau) = \frac{\langle I(t) I(t+\tau) \rangle}{\langle I(t) \rangle^2} = \frac{\mathbf{1}^\top \mathbf{V}_{\mathrm{D}} e^{[\mathcal{L}\mathbf{I} + \mathbf{K}_0]\tau} \mathbf{V}_{\mathrm{D}} \mathbf{p}_{\mathrm{ss}}}{(\mathbf{1}^\top V_{\mathrm{D}} \mathbf{p}_{\mathrm{ss}})^2} \tag{2.6}$$

where $\mathbf{V}_{\mathrm{D}}$ is the matrix of the monitored transition rates and $\mathcal{L}(r)$ the diffusion operator.[193] Similarly, FRET efficiency histograms and anisotropy decays can be computed from such simulations and compared to experiments.[171,201–203]

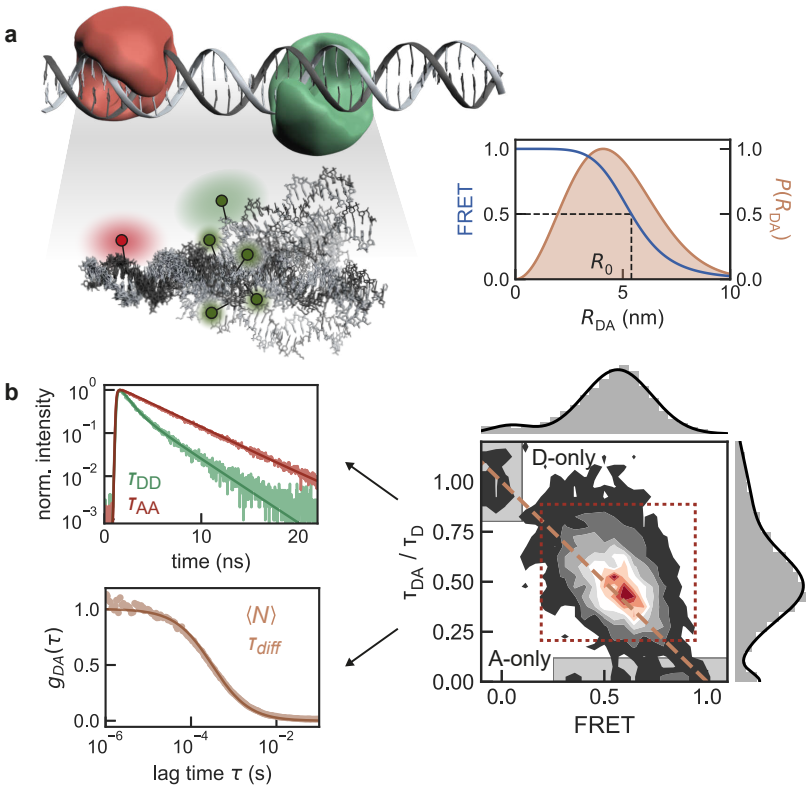


Fig. 2.3 | **Quantifying distance dynamics by single-molecule FRET**
a Double-stranded DNA labeled with a donor (green) and acceptor dye (red). The fluorophores diffuse within their sterically accessible volumes (colored cloud). The bending motions of the DNA helix and induce small variations in the donor-acceptor distance R_{DA} which can be expressed in the simplest case by a Gaussian chain model with a mean-squared end-to-end distance $\langle r^2 \rangle$ (Equation 2.5). **b** 2D histogram of the relative donor lifetime $\tau_{\mathrm{DA}}/\tau_{\mathrm{D}}$ against the FRET efficiency. Donor-only and acceptor-only populations are indicated. The double labeled species (red square) falls slightly above the static FRET line (orange, Equation 2.2). The donor-acceptor cross-correlation function reports on the translational diffusion time which is related to the hydrodynamic radius of the molecule by the Stokes-Einstein equation.

2.4 The future of FRET – challenges and perspectives

While the potential of FRET as a microscopic ruler has been recognized since the late 60s[215], it is the combination with single-molecule spectroscopy at the close of the century that brought FRET to the forefront of biophysical research. Since the first single-molecule FRET measurements in 1996[216], the method has allowed unprecedented insights into the conformational dynamics of biomolecules.[176,204,217–219] Its unique ability to resolve interconverting states and reaction intermediates at equilibrium positions single-molecule FRET as a powerful complement to structural biology techniques like X-ray crystallography, single-particle cryo-EM or nuclear magnetic resonance (NMR). These high-resolution methods capture snapshots of biomolecules at atomic detail in either crystallized, vitrified or solution form. However, often only a subset of all functional states are structurally accessible by chemical trapping. Intrinsic flexibility prevents many macromolecules from crystallizing. Ensemble averaging may result in an overlay of multiple conformations which results in broadening of resonances or smearing of electron density. In such cases, building a distance network by single-molecule FRET may resolve previously intractable conformers.[204,220] Integrative modeling which exploits experimental restraints from various sources is finding its way into structural biology.[218,221] Most recent developments include EM density restraints for *de novo* RNA structure prediction[162], the combination of electron paramagnetic resonance (EPR) and FRET-screening to explore transient conformational states inaccessible to crystallography[222], or distance restraints from cross-linking with mass spectrometry (XL-MS) to map protein-protein interactions.[223,224]

The main challenge in using FRET as a nano-positioning tool is the selection, production and measurement of a sufficiently large number of minimally redundant FRET pairs. For RNA a major bottleneck is sample production, namely the attachment of fluorophores at informative sites that do not interfere with the molecule's function. This often requires a tradeoff between what is considered to be the most relevant distance coordinates, and those sites that are functionally compatible or easily accessible (e.g. the termini). As long as the RNA is short enough to be chemically synthesized, the labeling chemistry can be tailored to the site of interest. Specifically, phosphoramidite building blocks can be decorated by different functional groups at various positions, be it on

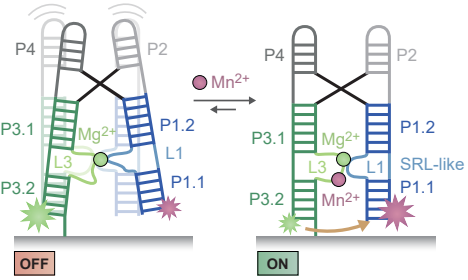
the sugar or the base, such that base-pairing remains unaffected. Transferring a functional moiety after transcription in a selective manner is much more difficult. To map large, non-coding RNAs by single-molecule FRET, we therefore need better labeling and purification strategies. A first step towards this goal is presented in chapter 3. Yet, there are still unaddressed issues. Among those are the efficient separation of labeled and unlabeled species or the need for novel bioorthogonal reactions which avoid modifications at the Watson-Crick face.

The second application area of single-molecule FRET leverages the high time resolution of fluorescence. It aims to unravel kinetic pathways with a limited number of donor and acceptor pairs. Reaction coordinates need to be selected such that individual states segregate well enough in time and space. More often than not, several conformational states may be inseparable based on distance alone. Let us consider a recurring RNA structural motif like the tetraloop/tetraloop receptor (TL/TLR) interaction. This non-WC interaction is frequently stabilized by divalent metal ions.[225,226] The structural response of the tetraloop-receptor towards Mg^{2+} binding may simply be too subtle for a detectable change in FRET efficiency. However, the dwell time of the bound state may dramatically increase as the contact gets stabilized by the ion. Since ion binding is transient though, short and long-lived states coexist, leading to heterogeneous kinetics. In combination with Markov modeling state connectivities could be successfully inferred for various such systems despite heavily overlapping FRET efficiencies.[79,177,178,182] To date, most studies on RNA have outlined kinetic networks that reduce to a classical four-state cycle.[74,79,80,133] A quantitative interpretation of more complex models has only rarely been put forward.[123,222] As a matter of fact, the number of possible pathways grows rapidly with each additional node in the graph. To unambiguously assign these states the dimensionality of the reaction coordinate needs to be high, hence many distances need to be measured. First initiatives to assess the precision and accuracy of distance measurements by FRET have been taken.[191,218] These multi-laboratory, blind challenges evaluate the current limitations in the field and provide benchmarks for better reproducibility.

RNA under the single-molecule microscope

The inherent flexibility of RNA molecules coupled with their propensity to partition into kinetically disperse subpopulations makes them ideal candidates for investigation by single-molecule spectroscopy. Our understanding of nearly every type of non-coding RNA has profited from advances in single-molecule detection: from those involved in gene expression (ribosome, tRNA), gene regulation (riboswitches, miRNA, siRNA) and immunity (CRISPR-Cas) to those RNA that are responsible for processing (snoRNA, telomerase) and catalysis (ribozymes, group I/II introns, spliceosome).[31] The following two examples illustrate this fruitful collaboration of two fields.

The **mechanism of signal transduction** in the *yybP-jkoY* RNA motif has recently been elucidated by a combination of X-ray crystallography and single-molecule FRET.[47] A single Mn^{2+} ion acts as a linchpin to secure the two coaxial stems of the riboswitch in a H-shaped geometry. Binding of the Mn^{2+} ion is assisted by a Mg^{2+} ion which pre-organize the binding site through formation of an A-minor interaction. A sarcin-ricin loop (SRL)-like motif in L1 is suggested to function as the molecular switch that prevents formation of the terminator stem and allows for transcriptional read-through (see also Fig. 1.6a).



RNA under the single-molecule microscope (continued)

Another recent contribution was made in the field of translation. Here, the **timing of transfer RNA (tRNA) selection** and dissociation of the elongation factor EF-Tu during proofreading has been revisited by three-color FRET and molecular dynamics.[197] Accommodation of cognate aminoacyl-tRNAs in the A-site of the ribosome and rejection of non-cognate tRNAs is shown to be temporally coordinated with the release of EF-Tu. Interestingly, in situations where peptide bond formation is slowed, EF-Tu was found to repetitively re-engage with the ribosome after guanosine triphosphate (GTP)-hydrolysis has occurred. Such increased turnover may improve the fidelity of translation at the cost of additional energy. Futile cycles of ternary complex formation (EF-Tu·tRNA·GTP) between with associated spurious consumption of GTP may also play a role in cellular responses to stress.

Chapter 3

Stick, Flick, Click: DNA-guided fluorescent labeling of long RNA for single-molecule FRET

Fabio D. Steffen*, Richard Börner, Eva Freisinger, and Roland K. O. Sigel*

Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

Abstract: Exploring the spatiotemporal dynamics of biomolecules on a single-molecule level requires innovative ways to make them spectroscopically visible. Fluorescence resonance energy transfer (FRET) uses a pair of organic dyes as reporters to measure distances along a predefined biomolecular reaction coordinate. For this nanoscopic ruler to work, the fluorescent labels need to be coupled onto the molecule of interest in a bioorthogonal and site-selective manner. Tagging large non-coding RNAs with single-nucleotide precision is an open challenge. Here we summarize current strategies in labeling riboswitches and ribozymes for fluorescence spectroscopy and FRET in particular. A special focus lies on our recently developed, DNA-guided approach that inserts two fluorophores through a stepwise process of templated functionality transfer and click chemistry.

Published in: *Chimia* **2019**, *73*, 257–261.
doi: 10.2533/chimia.2019.257

3.1 The ideal labeling strategy for RNA

Tags and labels are ubiquitous in nature. Post-transcriptional and post-translational modifications (PTM) provide the cell with a rich chemical repertoire that generates functional diversity within the transcriptome and proteome. These decorations are critically involved in signaling (phosphorylation by kinases), epigenetics (N-acetylation of lysine), they can change expression patterns (methylation of adenine, m6A), alter the cellular address of proteins and mark them for degradation (ubiquitylation), or report on their current folding status (glycosylation).[227,228] Like natural PTMs, extrinsic labels such as organic dyes, spin-labels or affinity tags can be incorporated into biomolecules both in vitro and in vivo to study their function or localize them in the cell.[229,230] For FRET and EPR applications, the challenge consists in making the modification specific to a unique site.[176,231,232] Compared to proteins, the alphabet of nucleic acid building blocks is more restricted and thus achieving site-selectivity becomes a major issue. The ideal RNA labeling method for fluorescence spectroscopy (1) can target any nucleotide within a biomolecule of any size; (2) it is site-specific down to a single nucleotide; (3) it is efficient, meaning the reaction is fast and gives a high yield; (4) it is bioorthogonal and proceeds without dye cross-reactivity, i.e. the reaction uses abiotic chemistry that is specific for each individual label; and (5) the method perturbs neither the RNA structure nor its function, thus the label should be small and not interfere with base-pairing.
A generally applicable method that would fulfill all the above criteria is lacking to date. Current labeling strategies for RNA suffer from the incompatibility between having full control over the labeling position and integrity of the RNA on the one hand, and the size of the biomolecule on the other (Fig. 3.1). Purely chemical approaches such as solid-phase oligonucleotide synthesis (SPOS) use modified phosphoramidites and are the method of choice for short sequences with < 60 nucleotides.[233,234] As such, SPOS gives the highest amount of flexibility in choosing the right dye position. To increase the size range of the target molecule, SPOS is often combined with enzymatic ligation where short chemically synthesized or in vitro transcribed fragments are combined into a single construct.[179,235] Using this approach, it has been possible to label riboswitches of up to about 200 nucleotides in length.[236,237] Still, the efficiency of the reaction depends on whether the secondary structure around the ligation site is tolerated by the enzyme. As an alternative

to T4 DNA and RNA ligase, Silverman and coworkers have identified deoxyribozymes through in vitro selection that ligate a fluorescently labeled (oligo)nucleotide via a 2′,5′-phosphodiester bond to an RNA of interest.[238,239]

A different approach learns from nature by imitating the functionality transfer reactions of natural PTM catalyzing enzymes.[240] A particularly attractive class are methyltransferases (MTase), which can be repurposed to deliver an S-adenosylmethionine (SAM) analog to the target RNA. Some of these promiscuous MTases are directed towards the 3′-terminus[241] or the 5′-cap[242] whereas others are programmable by a guide-DNA.[243] For live-cell imaging, chemo-enzymatic labeling emerges as a valuable alternative to genetically encoded RNA-fluorophore complexes like the spinach aptamer.[244]

On the other end of the spectrum are hybridization methods where short dye labeled DNA or peptide nucleic acids (PNA) oligonucleotides are attached to complementary regions on the RNA.[246,247] Usage of such fluorescent probes makes these strategies RNA size independent, at the expense of having a non-covalent and thus non-permanent linkage between the label and the target. Moreover, non-native loops often need to be inserted or elongated to accommodate the DNA or PNA oligo. Care must be taken not to disrupt key long-range tertiary contacts that would abrogate RNA function.

Recently, attempts have been made to combine the advantages of functionality transfer with the guiding capabilities of DNA-templated chemistry.[248–250] To label particularly long non-coding RNAs like riboswitches and ribozymes, we developed a novel site-directed labeling approach, which installs two fluorophores in a post-transcriptional and bioorthogonal manner.[245] In the following, we will review the conceptual basis of the labeling workflow with a special focus on the chemistry of the transfer and coupling reactions.

3.2 Guide, transfer, couple – RNA labeling in three steps

Egloff et al. recently described an alkylation strategy to site-specifically incorporate etheno adducts on the Watson-Crick face of adenines and cytosines in single-stranded DNA oligonucleotides.[248] If the alkylating agent is decorated with a bioorthogonal functional group, this moiety can be subsequently conjugated with a fluorophore or any other reactive probe.[251] The applicability of this two-step approach has been demon-

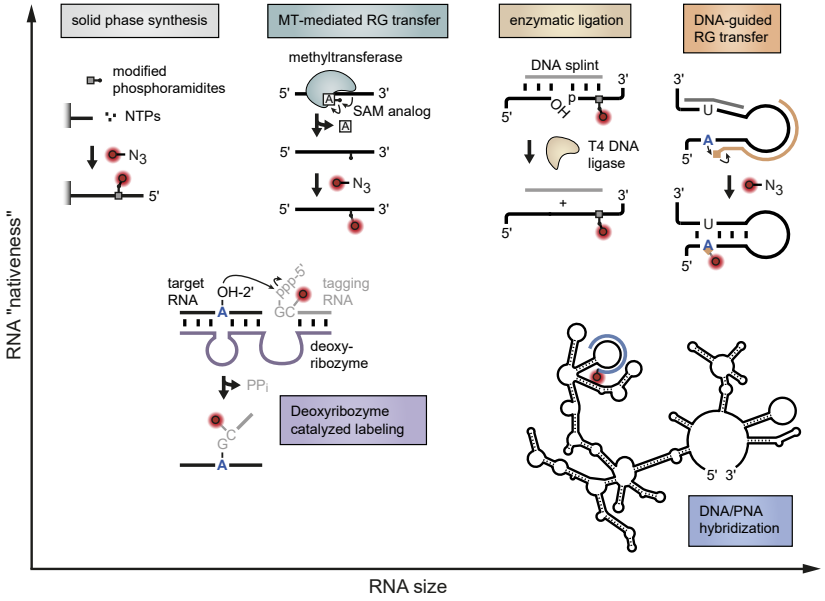


Fig. 3.1 | **Strategies for site-specific RNA labeling.** Most current labeling methods are limited by the RNA size or its "nativeness" (i.e. integrity of sequence and structure). Solid-phase oligonucleotide synthesis (SPOS): the RNA is chemically synthesized using phosphoramidite building blocks and labeled post-synthetically (e.g. by click chemistry).[233,234] Methyltransferase (MT) mediated reactive group (RG) transfer: the enzyme is promiscuous towards S-adenosyl methionine (SAM) analogs and can transfer a functional group (e.g. an alkyne) instead of the canonical methyl group.[240] Deoxyribozyme catalyzed labeling (DECAL): an in vitro selected deoxyribozyme couples a fluorescently labeled tagging RNA onto the target RNA.[238] The tagging RNA can be as short as a single guanine.[239] Enzymatic ligation: synthesized or transcribed oligonucleotides are enzymatically ligated with a DNA splint.[179,235] DNA-guided RG transfer: a bioorthogonal handle is guided towards a pre-selected adenine and subsequently clicked with a dye.[245] DNA/PNA hybridization: fluorescently labeled DNA or PNA oligonucleotides anneal non-covalently to engineered loops platforms.[246,247]

strated on a surface exposed single-stranded loop of a several hundred nucleotide long group II intron ribozyme.[251] To extend the scope of labeling positions, we introduced additional helper strands[252] which anneal upstream and/or downstream of the modification site and temporarily disrupt the local secondary structure pattern, making the selected nucleotide better accessible for the chemical modification (Fig. 3.2a).[245] The labeling sites are thus no longer limited to single-stranded regions but also base-paired nucleotides can be targeted. The functionalized guide-DNA then brings a vicinal diol into proximity of the exocyclic amine group of an adenine or cytosine. The diol is cleaved by periodate leaving an aldehyde to react with the nucleophilic primary amine. In a cyclization reaction, followed by elimination of water, a 1,N6-ethenoadenine (ϵA) or 3,N4-ethenocytosine (ϵC) is generated (Fig. 3.2b). The propargyl handle on position C8 is subsequently coupled to an azide derivatized fluorophore using Cu(I) catalyzed click chemistry (CuAAC). As a useful side-reaction of the in situ activation of the diol, the 3′-terminal ribose is oxidized and can be targeted in an orthogonal fashion using a second dye.[253,254]

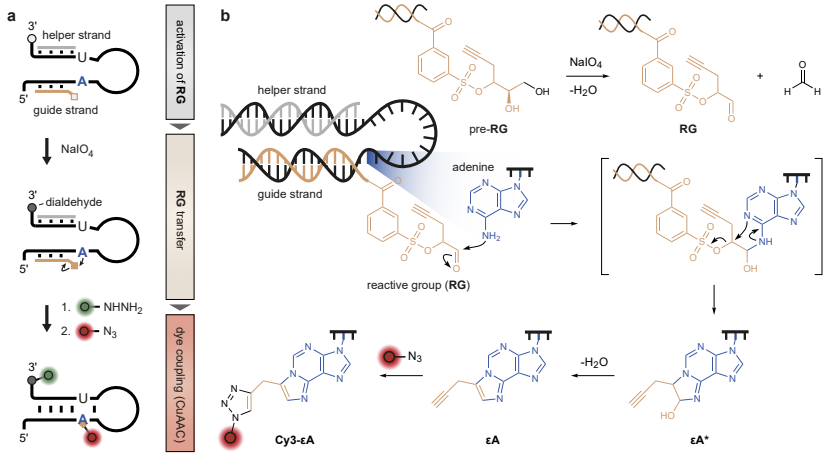


Fig. 3.2 | **DNA-guided fluorescence labeling of RNA in three steps.** (a) Schematic overview of reactive group activation, transfer and bioorthogonal dye coupling. (b) Helper (gray) and guide (yellow) strands hybridize to the target RNA to bring a reactive group (RG) into the reach of a pre-selected adenine (blue). $NaIO_4$ activates the RG by oxidative cleavage. Concomitantly, the 3′-terminal ribose is converted to a dialdehyde. Etheno adduct formation proceeds through a hydroxyethano intermediate (ϵA*). In the last step, an azide functionalized dye is conjugated to the alkyne in a Cu(I) catalyzed cycloaddition reaction (CuAAC). A second, hydrazide derivatized fluorophore reacts with the 3′-terminal dialdehyde.

3.3 Site selection – where to put the label?

Single-molecule detection combined with Förster resonance energy transfer (FRET) has evolved into a versatile tool to monitor biomolecular interactions and dynamics on a broad range of timescales from nanoseconds to minutes or even hours.[175,255] The selection of appropriate dye positions is thereby a prerequisite for informative FRET trajectories. To learn about RNA dynamics (e.g. a conformational rearrangement initiated upon binding of a metabolite or a protein), the dye pair should probe a functionally relevant reaction coordinate. In a two-state folding setting this can be two residues which are far apart in the unfolded state (low FRET) but come close together upon RNA collapse (high FRET). To use FRET as a spectroscopic ruler with maximum sensitivity,[211,215] the expected interdye distance r should match the Förster radius R_0R of the FRET pair (Eqn. 3.1). Furthermore, reliable transfer efficiencies E require the dyes to rotate isotropically, which relates to the well-known "κ^2-problem".[200] Most importantly, the fluorophore positions have to be compatible with the RNA structure and function.

$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \tag{3.1}$$

Riboswitches are regulatory RNAs that undergo a conformational change in response to binding of small-molecule metabolites such as S-adenosylmethionine (SAM),[36] flavin mononucleotide (FMN)[37] or cobalamins (e.g. coenzyme B_{12}).[38,256] Hence, riboswitches represent a prime example of biomolecular dynamics that can be probed by single-molecule FRET.[74,79,80] To showcase the applicability of our labeling protocol, we use a coenzyme B_{12} responsive riboswitch that controls the expression of an outer-membrane protein responsible for the transport of coenzyme B_{12} in Escherichia coli.[60]

To guide our search for appropriate labeling positions, we based our selection on a consensus secondary structure from a multisequence alignment (MSA) of cobalamin riboswitches[66] as well as a crystal structure of one representative.[86] We selected an adenine close to a kissing loop contact[66] (A35) and another that is part of the ribosomal binding site (RBS) hairpin[62] (A213). Both residues are thus near key functional elements but not themselves involved in long-range tertiary contacts. The two internal sites are complemented by the 3′-terminus (G240) which gets modified concomitantly. In this way, FRET monitors the formation

of the kissing loop and the RBS hairpin by probing a reaction coordinate from the donor labeled G240 to the acceptor labeled A35 or A213, respectively.

3.4 Quality checks – site-specificity, dye integrity and RNA switching

To test the precision of our DNA-guided functionality transfer reaction, we designed a reverse transcriptase assay, where a short ^{32}P-carrying primer is annealed downstream of the labeling site (here A35) and is then stepwise extended by a polymerase (Fig. 3.3a). On a denaturing PAGE, bands appear for every nucleotide that is attached to the growing primer, except for the one where we expect the fluorophore to be conjugated to. The absence of this band suggests that the reverse transcriptase does not recognize the labeled nucleotide, as its Watson-Crick face is blocked, and therefore does not pause and release its cDNA product. This shows that out of two consecutive adenines (A35 and A36) primarily A35 is labeled.

Fluorescence is ideally suited to interrogate dynamic processes in biomolecular settings because it is non-invasive and can cover a wide range of timescale from dye photophysics and diffusion (picoseconds to microseconds) to folding and binding kinetics (microseconds to minutes).[175,255] The fluorophores should thus be bright (i.e. have a high quantum efficiency), they should be photostable, water soluble and offer biocompatible coupling chemistries.[257] One popular class of dyes for single-molecule FRET are carbocyanines. Their photophysics and interaction profiles with nucleic acids and proteins have been extensively characterized.[167,170,171] Long flexible linkers between the attachment point and the dye scaffold, as well as negatively charged sulfo groups help to minimize contacts with the biomolecule and enhance the free rotation of the dye.[258] Nevertheless, interactions particularly at the termini are not uncommon.[259,260] Consistent with previous fluorescence lifetime and time-resolved anisotropy measurements,[171] the typical long fluorescence decay curves indicate that the dyes are still intact after conjugation to the RNA even if their flexibility is restrained by the biomolecule (Fig. 3.3b).

To evaluate the integrity and function of the riboswitch, we immobilize the RNA at the 5′-terminus via a streptavidin-biotin linker onto coated

quartz slides and probe the RBS hairpin by single-molecule FRET.[261] In order to distinguish zero FRET molecules (inter-dye distance $> 10\,\text{nm}$) from donor only or acceptor bleached molecules, we additionally check the presence of the acceptor dye with a red laser in an alternating laser excitation (ALEX) scheme.[262] In this way, we can safely select only those molecules that are double labeled.
The single-molecule time traces and histograms suggest that the RBS exists in at least two major conformations, an open (low FRET) and a closed (high FRET) state (Fig. 3.3c). The open conformation possibly encompasses an ensemble of structures where the RBS is accessible for the ribosome to bind and translate the downstream gene. In the closed form, the RBS is sequestered and protein expression is downregulated.[62] Binding of the natural cofactor coenzyme B_{12} (CoB_{12}) to the aptamer region shifts the equilibrium towards the closed state, consistent with a negative feedback mechanism.[256] By observing an increasing population of the high FRET state upon metabolite binding, we are confident that introduction of the bulky fluorophores at the designated positions does not impair the function of the riboswitch.

3.5 Multi-colored RNA labeling – limits and prospects

We review here a recently established two-color labeling method which is applicable to RNAs of any size while preserving its native structure and sequence as good as currently possible. The labeling is precise and adaptable to different spectroscopic techniques such as FRET or EPR. With respect to the ideal labeling strategy there are still a few shortcomings: (i) Because ethenoadduct formation involves nitrogen atoms N1 and N6 of adenine or N4 and N3 of cytosine, the Watson- Crick base pairing of the modified residue is disrupted. Thermal melting experiments show that effects on the stability of longer RNAs are small, suggesting that only the tagged residue is affected. Special care is advised when targeting a functionally relevant, short and AU rich duplex. (ii) Double labeling yields are around 5% (15–35% for a single dye), which is usually enough for single-molecule measurements but may be a constraint for ensemble experiments. Optimization of the dye coupling stoichiometry and overall RNA recovery during purification might alleviate this issue. (iii) Separation of labeled from unreacted RNA is challenging. Size differences are small (usually $< 1\,\text{kDa}$) and the RNA is not significantly more hydrophobic due to the presence of the dyes alone, since common

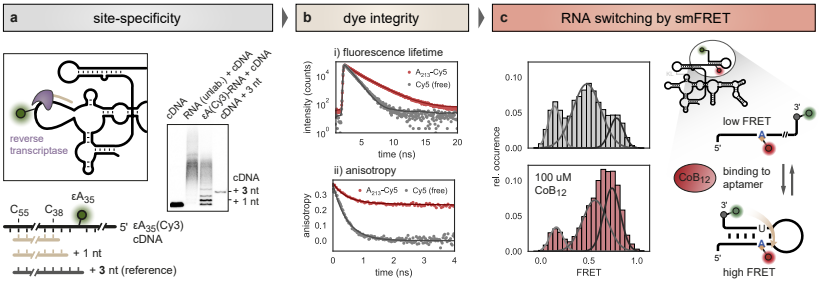


Fig. 3.3 | **Quality checks to assess the site-specificity, fluorophore integrity and RNA switching.** (a) Reverse transcriptase assay probing the internal labeling of a B_{12} responsive riboswitch with Cy3 on A35. The denaturing PAGE gel shows a ^{32}P-labeled cDNA primer which gets extended by a reverse transcriptase. The band corresponding to the cDNA primer + 3 nucleotides, which is opposite to the labeling site, is missing because the reverse transcriptase does not recognize ϵA35-Cy3 with its Watson-Crick face being blocked. The enzyme therefore does not pause at his position and no cDNA product is released. Gel reproduced from ref. [245] with permission from Oxford University Press. (b) Prolonged fluorescence lifetime and time-resolved anisotropy decays indicate that the fluorophores are conjugated to the RNA and are still intact. Decays adapted from ref. [245]. (c) Riboswitch dynamics interrogated by single-molecule FRET. Binding of coenzyme B_{12} (CoB_{12}) to the RNA aptamer region triggers a switch on the expression platform, pushing the ribosomal binding site (RBS) hairpin into a closed form (high FRET), which in turn downregulates gene expression. Histograms adapted from ref. [245].

fluorophores are often sulfonated to minimize interactions. Unlabeled or monolabeled riboswitches will compete for metabolites, but those species can be sorted out by dual-color excitation in single-molecule imaging. Ultimately, the key advantage over other existing approaches is the independence of RNA length, while establishing a covalent and thus permanent linkage between the RNA and the dye. As such, the method combines the main benefit of DNA/PNA hybridization with direct dye coupling as in solid-phase oligonucleotide synthesis. The current design of the reactive group has been tested to tag adenines and cytosines but should be extendable to guanines as well.[263] Furthermore, alternative transfer chemistries at different positions on the base or sugar ring are conceivable. PNA instead of DNA-guided delivery of the RG would allow to shorten the guide strand and possibly increase labeling yields at highly buried residues. Lastly, the reactive group chemistry could be reversed (i.e. conjugate a RG-azide with a dye-alkyne) to enable orthogonal dye couplings at two or more internal sites. To avoid the cellular toxicity of Cu(I), other bioorthogonal reactions (e.g. copper-free click chemistry with strained cyclooctynes,[264] Staudinger ligation[265] or inverse-electron demand Diels-Alder reactions[266]) could be exploited in the future. All these developments build upon the modular nature of the method (DNA-guided delivery – stick, RG transfer – flick, and dye coupling – click) and further expand the scope of possible labeling sites within biologically relevant RNAs.

3.6 Acknowledgments

The authors thank Meng Zhao for many inspiring discussions on RNA labeling. This work was supported by the Swiss National Science Foundation (to EF and RKOS), SystemsX.ch (to RKOS) and the UZH Forschungskredit (FK-17-098 to FDS; FK-14-096 and FK-15-095 to RB).

Chapter 4

Metal ions and sugar puckering balance single-molecule kinetic heterogeneity in RNA and DNA tertiary contacts

Fabio D. Steffen, Mokrane Khier, Danny Kowerko, Richard A. Cunha, Richard Börner*, Roland K. O. Sigel*

Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

Abstract: The fidelity of group II intron self-splicing and retrohoming relies on long-range tertiary interactions between the intron and its flanking exons. By single-molecule FRET, we explore the binding kinetics of the most important, structurally conserved contact, the exon and intron binding site 1 (EBS1/IBS1). A comparison of RNA-RNA and RNA-DNA hybrid contacts identifies transient metal ion binding as a major source of kinetic heterogeneity which typically appears in the form of degenerate FRET states. Molecular dynamics simulations suggest a structural link between heterogeneity and the sugar conformation at the exon-intron binding interface. While Mg^{2+} ions lock the exon in place and give rise to long dwell times in the exon bound FRET state, sugar puckering alleviates this structural rigidity and likely promotes exon release. The interplay of sugar puckering and metal ion coordination may be an important mechanism to balance binding affinities of RNA and DNA interactions in general.

Published in: *Nature Communications* **2020**, *11*, 104.
doi: 10.1038/s41467-019-13683-4

4.1 Introduction

In the course of RNA maturation, self-splicing ribozymes catalyze two consecutive reactions, that is, the excision of the intron and the ligation of its flanking exons.[267] To precisely locate the cleavage site, group I and II introns engage specific RNA recognition elements that base pair with complementary stretches on the exons (Fig. 4.1a).[268–270] In group II introns, these sequences are not conserved, instead the RNA identifies the structural transition between single- and double-stranded bases as the 5′-splice site.[271] The long-range tertiary interactions embed the 5′-exon in the active core of the ribozyme (Fig. 4.1b)[272–274], where a hydrogen-bond network and several coordinating metal ions convey stability to the tertiary contacts.[114,275] Two Mg^{2+} directly participate in catalysis by activating the scissile bond for cleavage.[276] Ideally, the metal ions fine-tune the interaction such that the 5′-exon is held in place until the second transesterification has occurred and the two exons are ligated. If the interaction between intron and exon is too weak, splicing either does not occur at all (intron retention) or stops after the first step, with the 5′-exon leaving the active site without being ligated. Conversely, if the association of intron and exon is too strong, the ribozyme no longer discriminates between correct and mismatched targets, which may lead to gene disruption and disease if reverse splicing occurs in tumor suppressor genes like p53.[277,278] To minimize such errors, most group II introns use two independent exon recognition sites to keep hold of the 5′-exon: exon binding site 1 (EBS1) contributes the most to thermodynamic stability, while the surface-exposed EBS2 is responsible for target selectivity.[278,279] Mono- and divalent ions play their part in strengthening the intron-exon interaction both through non-specific charge screening as well as site-directed coordination. A number of such specific binding sites have been identified near active site elements.[114,275] Like a padlock, the metal ion packs the strands together, thereby retaining the exon in the active site for a longer time than if no gatekeeping ion was present. As ion coordination is transient though, designated binding pockets are often only partially occupied and exon dissociation is thus kinetically heterogeneous.[280] Particularly, Mg^{2+} is known to induce such kinetic partitioning by interacting with RNA directly (inner-sphere coordination) or via a water molecule (outer-sphere coordination).[94,108,175] Single-molecule techniques are well suited to detect subpopulations of molecules in different conformational states.[80,226,280,281] This involves

ion-induced collapse of RNA secondary structure elements as well as formation of more distant tertiary contacts.[108] Spectroscopic rulers such as Förster resonance energy transfer (FRET) capture these rearrangements based on a set of predefined distance coordinates. Yet, in some cases, be it due to the choice of the dye positions or the local geometry of the coordination environment, metal ion binding yields no detectable change in resonance energy transfer and results in a degenerate FRET state that comprises multiple kinetic states. Their existence is only inferred from the multiexponential decay rates from the one FRET state. As metal ions direct folding and catalysis, heterogeneity is inherent to many RNAs.[175,282] RNA folds hierarchically into a set of interconnected topological modules. The kinetics of their tertiary assembly is often rate determining in the folding process.[108] Pseudoknots are a recurrent motif in riboswitches, ribozymes, and the ribosome where they interface secondary and tertiary structure. They are often involved in the formation of the catalytic core or the aptamer domain[78,129,283,284]. In its simplest form, an H-type pseudoknot consists of a hairpin that base pairs intramolecularly with a stretch of nucleotides outside the stem.[285] In group II introns such a hairpin reaches out to the flanking exon and forms the recognition site for splicing and retrohoming.[268,286–288] Here we use single-molecule FRET in combination with hidden Markov modeling to probe the pairing of EBS1 in the prototypical group II intron *Sc*.ai5γ with its cognate intron binding site 1 (IBS1). To facilitate the interpretation of the binding free energy landscape, we have isolated the tertiary contact and tagged both interaction partner covalently with a fluorescent dye. In this way, we establish FRET as a reaction coordinate, which directly reports on the binding of exon and intron. Formation of this contact is a prerequisite for subsequent catalysis. While structures of the exon-intron complex before the first step of forward and reverse splicing have been solved recently[273,289,290], dynamic information on the association and dissociation reaction is still largely missing to date. By comparing how RNA and DNA exons bind to the isolated EBS1 hairpin, we can dissect the kinetics of the recognition step, which precedes forward and reverse splicing. We find that RNA-DNA hybrids not only dissociate much faster than their pure RNA counterparts, but they do so in a more kinetically homogeneous manner. Our molecular dynamics simulations show that the origin lies in the conformational fit between the binding partners. Fast switching of the sugar puckers in the DNA exon compromises tertiary contact stability and likely impairs metal ion

binding. This smooths out the rugged energy landscape and homogenizes the kinetics.[107] Hence, sugar puckering is an effective way to counteract ion-induced heterogeneity in exon recognition and is a potentially widespread mechanism to modulate macromolecular interactions.

4.2 Results

4.2.1 DNA target recognition requires Mg^{2+}

We monitor the binding and unbinding of two seven nucleotides long exon strands, IBS1* (RNA) and dIBS1* (DNA), to surface-immobilized EBS1* hairpins by following the anti-correlated Cy3 and Cy5 emission over several minutes (Fig. 4.1c and Supplementary Fig. 4.6). The two intensity signals are converted into transfer efficiencies that fluctuate between a zero FRET state, corresponding to the unbound hairpin, and a high FRET state around 0.75 of the formed tertiary contact. We classified the molecules according to their interconversion frequency into (i) static zero (no binding event within observation time), (ii) static high (no unbinding event), (iii) one transition and (iv) dynamic (at least two transitions). The relative abundance of all four classes across different Mg^{2+} concentrations is summarized in Supplementary Fig. 4.7. Both RNA-RNA and RNA-DNA contacts are remarkably sensitive to Mg^{2+}, as indicated by the strong depletion of static zero FRET molecules in favor of dynamic molecules when Mg^{2+} is added (Fig. 4.1d). Static traces in the high FRET state only appear in the presence of IBS1*, but not dIBS1* (Fig. 4.1e). These persistent contacts (>400 s) are unique to the RNA-RNA contact and are strictly dependent on Mg^{2+} (Supplementary Fig. 4.7). This suggests that Mg^{2+} not only condenses around the RNA but binds site specifically to the tertiary contact. Further evidence for such a binding site comes from nuclear magnetic resonance (NMR) chemical shift mappings with Mg^{2+} and $[Co(NH_3)_6]^{3+}$, as well as Mn^{2+}-induced line broadening.[291,292] The affected nucleotides coincide with a patch of negative surface potential near the 5′-splice site. Upon docking of IBS1* or dIBS1*, a cavity is formed, which readily accommodates a Mg^{2+} ion.[291,292] Divalent ions are therefore expected to promote the interaction of exon and intron by a combination of non-specific charge compensation and site-specific binding.[92,94] The fraction of formed contacts are deduced from FRET efficiency histograms of all dynamic molecules (Fig. 4.1f). The relative occupancy of the high FRET state increases as

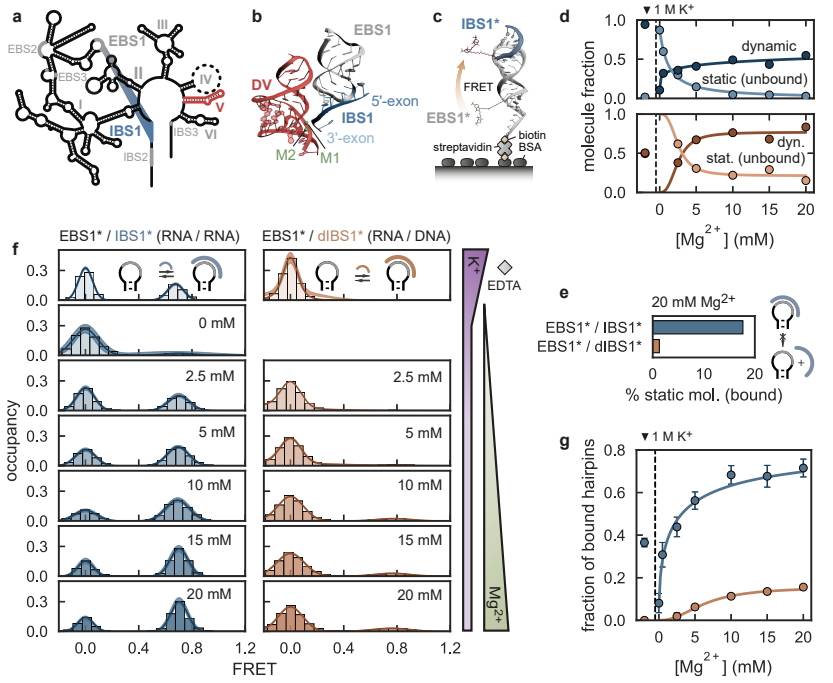


Fig. 4.1 | **Thermodynamics of Mg^{2+}-induced tertiary contact formation from single-molecule FRET. a** Secondary structure of the group IIB intron *Sc*.ai5γ with the exon-intron binding site 1 (EBS1/IBS1, gray/blue) and the catalytic domain V (red). **b** Crystal structure of the active site of the *P.li*.LSUI2 intron (PDB: 4R0D[274]) featuring domain V (DV, red), EBS1/IBS1 (gray/blue), and two catalytic metal ions (M1/M2, green). **c** Surface-immobilized FRET model system EBS1*/IBS1* labeled with Cy3 and Cy5 (BSA, streptavidin, and biotin are not drawn to scale). **d** Relative abundance of dynamic and statically unbound IBS1* (blue) or dIBS1* (orange) molecules as a function of Mg^{2+}. **e** Percentage of long-lived, statically bound molecules at 20 mM Mg^{2+} and 100 mM K^+. **f** FRET histograms of dynamic molecules showing a rising population of the high FRET state with increasing Mg^{2+} concentration. **g** Fraction of EBS1* hairpins bound to (d)IBS1* as a function of the Mg^{2+} concentration calculated from the integrals of the FRET states. Error bars correspond to the mean $\pm$ 2 s.d. of 100 bootstrap samples.

a function of the Mg^{2+} concentration (Fig. 4.1g). The normalized binding isotherm saturates at around 70% of bound EBS1*/IBS1* molecules. The fraction of bound RNA-DNA on the contrary does not exceed 20% even at 20 mM Mg^{2+}. Complete saturation of the hairpin would require exon concentrations in the high micromolar to millimolar range, which is inaccessible to single-molecule total internal reflection fluorescence (TIRF) imaging because of the high background signal (Supplementary Fig. 4.8). The dissociation constant calculated from the equilibrium population of the zero and high FRET states shows that EBS1*/dIBS1* is about an order of magnitude less stable than EBS1*/IBS1* (Supplementary Fig. 4.9a). This observation is consistent with previous reports of RNA-DNA hybrid duplexes being more labile than their canonical RNA-RNA counterparts.[293–295]. Notably, no EBS1*/dIBS1* formation at all is observed in the absence of Mg^{2+}, highlighting the importance of divalent ions in stabilizing tertiary structure motifs. The free energy contribution of Mg^{2+} binding to tertiary contact formation is given by the difference in Mg^{2+} binding to the hairpin alone and to the formed contact, involving both diffuse and site-specific components (Supplementary Information and Supplementary Fig. 4.9b). Diffuse binding is expected to be similar for RNA and DNA exons, thus the shift of the midpoint of the binding curve (RNA-RNA: $[Mg^{2+}]_{mid} = 3.7 \pm 0.9$ mM versus RNA-DNA: $[Mg^{2+}]_{mid} = 7.9 \pm 1.9$ mM) suggests a tighter inner and/or outer-sphere coordination of Mg^{2+} at the RNA-RNA interface in line with previous binding affinities calculated from NMR chemical shifts.[280,296]

4.2.2 Mg^{2+} induces heterogeneity by slowing down exon dissociation

Among the four molecular classes defined above, dynamic molecules are the most informative ones as they all represent functional RNA tertiary contacts that associate and dissociate within the observation window. To explore the kinetics of the interaction between the exon and intron binding sites, we discretized the FRET trajectories of the dynamic molecules into bound and unbound segments as shown in Fig. 4.2a. A visual inspection of the traces reveals both longer dwell times and more frequent transitions to the high FRET state, the more Mg^{2+} is present. By computing a mean dwell time $\langle t_{zero} \rangle$ and $\langle t_{high} \rangle$ for each trace, we can assess how uniformly the immobilized EBS1* hairpins react to IBS1* and dIBS1* binding (Fig. 4.2b). If there are kinetic subpopulations of EBS1*

molecules, some of which bind IBS1* stronger than others, we would expect them to separate into individual clusters in the dwell time scatter plot as suggested by simulations (Supplementary Fig. 4.10a). On the other hand, if short and long binding events occur within the same trace, they will average out into a single stretched cluster. The width of the distribution is further influenced by the number of dwell times and thus by the limited observation time (Supplementary Fig. 4.10b, c). There is indeed no clear partitioning of the molecules into separate clusters (low molecule/subspecies heterogeneity, Fig. 4.2b). We observe, however, a positive correlation between the mean dwell time and the variability within a trace, measured by the difference between the shortest and the longest dwell time (Δt_{zero} or Δt_{high}, Supplementary Fig. 4.11). It follows that the most heterogeneous molecules (large Δt) also feature long mean dwell times. To test whether these long dwell times are reinforced by divalent metal ions, we monitor the center of the mean dwell time ensemble across different Mg^{2+} concentrations (Fig. 4.2c and Supplementary Table 4.2). In the absence of Mg^{2+}, diffuse K^{+} ions stimulate strand association mainly through charge compensation, but keep the excursions to the bound state relatively short. On average, IBS1* falls off the hairpin after about 20 s, and dIBS1* unbinds almost immediately after docking (0.4 s). In contrast, addition of 20 mM Mg^{2+} ion triples the average residence time of the IBS1* exon and prolongates the dwell time of dIBS1* by a factor of 10. This goes along with an overall broadening of the distribution of mean dwell times (Supplementary Fig. 4.12). Hence, we propose that kinetic heterogeneity is induced predominantly by specific binding of Mg^{2+} to preformed tertiary contacts, which slows down the exon dissociation rate. Because Mg^{2+} coordination occurs on a similar timescale as exon unbinding, as shown by previous Mg^{2+} pulse experiments[280], but is usually faster than the observation time, short- and long-lived dwell times coexist within the same trace.

4.2.3 Off-rates determine tertiary contact stability.

A kinetic comparison of the relatively stronger RNA-RNA and the weaker RNA-DNA interaction is useful in identifying the stability determining rate of tertiary contact formation. It can be viewed as a global ϕ-analysis, where instead of specific mutations, RNA is substituted with the same sequence of DNA to perturb the energetic levels of the bound native state, the transition state, or both.[297,298] We found that the

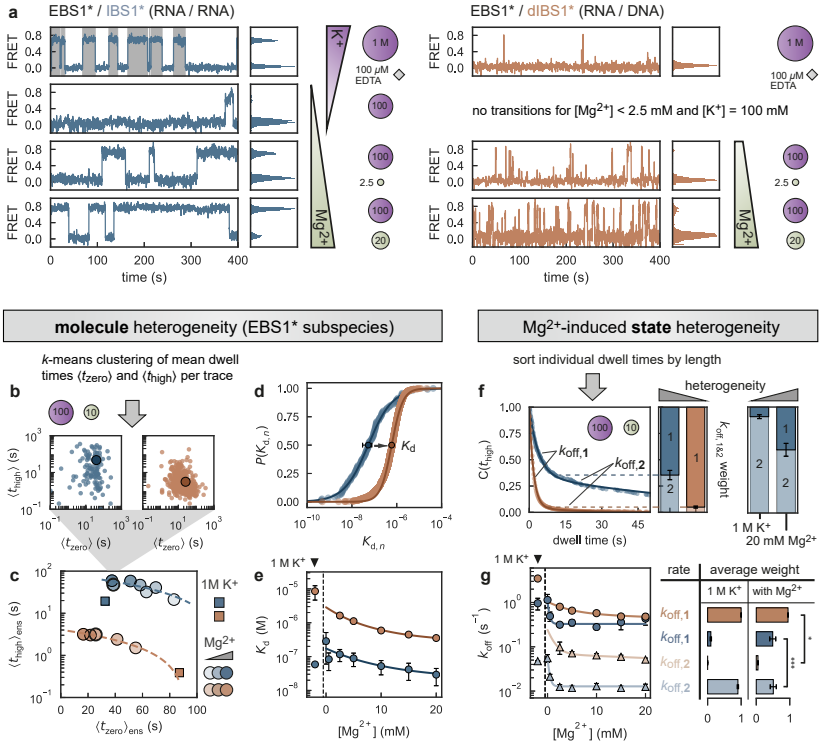


Fig. 4.2 | **Kinetic heterogeneity is tuned by Mg^{2+} and the nucleic acid strand type.**

a Representative single-molecule traces of EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange) displaying different residence times in the zero and high FRET state in response to gradual changes in the composition of the ionic environment. The traces are selected to reflect the overall histogram shown in Fig. 4.1f. The discretized state sequence is depicted by alternating backgrounds (white: zero FRET; gray: high FRET) for the top left trace and omitted elsewhere for clarity. Only dynamic molecules were used for the analysis. **b** The molecule-to-molecule variability is visualized by correlating the mean dwell times in the bound, $\langle t_{\mathrm{high}} \rangle$, and unbound state, $\langle t_{\mathrm{zero}} \rangle$, for every time trace. There are no distinct kinetic subspecies, but rather a single stretched cluster. Data is shown for 100 mM K^+ and 10 mM Mg^{2+}. **c** The cluster center is calculated by k-means and depicted as a circle in different shades of blue or orange for every Mg^{2+} concentration (10 mM Mg^{2+} is highlighted with a thick, black edge). Contact formation under high ionic strength (1 M K^+) in the absence of Mg^{2+} (100 µM EDTA) is indicated by a blue or orange square. (*continued on next page*)

Fig. 4.2 | (*continued*) **d** Cumulative $K_{\mathrm{d},n}$ histograms (10 mM Mg^{2+}) are fitted to a logistic function (Equation 4.2). **e** Stability constants of EBS1*/IBS1* and EBS1/dIBS1 as a function of Mg^{2+} are quantified by the transition point, K_{d}, from fits in c. Error bars are drawn as mean ± 2 s.d. of 100 bootstrap samples. **f** Heterogeneity within the bound state is evaluated by computing cumulative dwell time histograms of the bound state. The presence of multiple decays is a characteristic of kinetic heterogeneity. The weighting factors (mean ± s.d.) of the dissociation rate coefficients $k_{\mathrm{off},1}$ and $k_{\mathrm{off},2}$ are indicated alongside the decay. The ion dependence of the prefactor is displayed on the right for EBS1*/IBS1*. **g** Rate coefficients and average weighting factors (mean ± s.d.) over the whole Mg^{2+} titration. The star (*) indicates the degree of heterogeneity. The latter is largest if the rate coefficients are far apart and equally populated (weighting factor ∼0.5).

RNA-RNA and RNA-DNA differ mainly in the average time IBS1* or dIBS1* spends bound to the hairpin, that is, $\langle t_{\mathrm{high}} \rangle$. In contrast, the average unbound dwell $\langle t_{\mathrm{zero}} \rangle$ time is largely independent of the strand type. The ratio of unbound and bound mean dwell times translates into a dissociation constant, $K_{\mathrm{d},n}$, for each trace. The distribution of all $K_{\mathrm{d},n}$ is represented in a cumulative histogram and fitted to a logistic function (Equation 4.2, Fig. 4.2d, and Supplementary Fig. 4.13). The K_{d} at the inflection point is the expected value of the distribution and a measure for the stability of the tertiary contact. Over the Mg^{2+} titration (0 mM–20 mM) the K_{d} of EBS1*/IBS1* decreases by a factor of 10, resulting in an affinity of 29 ± 8 nM at 20 mM Mg^{2+} (Fig. 4.2e). The K_{d} of EBS1*/dIBS1* is about one order of magnitude higher under the same salt conditions (352 ± 8 nM at 20 mM Mg^{2+}), reflecting the lower stability of the RNA-DNA hybrid compared to the RNA-RNA, consistent with previous surface plasmon resonance (SPR) experiments (Supplementary Table 4.3).[292] The similar on-rates of the two contacts suggest a reactant-like transition state where the exon and intron are in proximity, but base pairs are not yet fully formed. The energetic barrier originates mostly from the entropic cost of freezing backbone motions as well as uptake and localization of metal ions in the transition state.[226,299] Because the energetic level of the transition state is invariant to the type of exon, the stability of the tertiary contact is dictated by the off-rates. These observations are in line with the emerging paradigm of early transition states being a hallmark of RNA folding.[129,226,295,297] On a rugged

energy landscape, it is the rate at which wrongly formed base pairs can be broken again, which limits the overall speed of folding.[108] Mg^{2+} ions provide both shortcuts along the folding route but also trap molecules in misfolded conformations. It is this dual role that leads to the stretched folding times characteristic for many ribozymes.[122,282]

4.2.4 RNA-DNA is less kinetically heterogeneous

So far, we have looked at the average dwell time of (d)IBS1* in its bound and unbound state. This analysis gives us a robust estimate of the tertiary contact stability from the ensemble of single molecules. On the downside, variations in dwell time length within one trace are averaged out. Mg^{2+} titrations suggest that these variations are caused by ion coordination. Mg^{2+} binding is subtle such that it does not alter the inter-fluorophore distance and consequently ion bound and unbound populations display the very same FRET efficiency. In other words, a single FRET state degenerates into two interconverting kinetic states, each with its own off-rate. We refer to this scenario as state heterogeneity. To determine the different rates, we computed complementary cumulative dwell time distributions of the bound and unbound state (Fig. 4.2f, Supplementary Fig. 4.14, and Supplementary Table 4.4). Binding kinetics are well described by a single exponential and a stretching factor β to account for small deviations at longer times. Unbinding, on the other hand, is not homogeneous, but partitions into two exponential terms in the presence of low millimolar amounts of Mg^{2+}. The relative weight of these two components differs considerably depending on the nature of the complementary strand. At 20 mM Mg^{2+}, the fast off-rate, $k_{off,1}$, accounts for 40 ± 6 % of the RNA-RNA contact decay, whereas its amplitude rises to 96 ± 1 % in the RNA-DNA contact, meaning the decay is practically homogeneous for the hybrid interaction, even though a small percentage of longer-lived contacts persists (Supplementary Fig. 4.14). In either case, off-rates are most sensitive to low millimolar Mg^{2+} concentrations (0 mM–2.5 mM), which coincides with the K_d of Mg^{2+} to the tertiary contact (Mg^{2+} binding in the loop: 1.78 ± 0.01 mM, binding at the stem-loop transition: 0.87 ± 0.01 mM, Fig. 4.2g and Supplementary Fig. 4.15).[280] A comparison of the IBS1* and dIBS1* dissociation rates reveals that the higher stability of EBS1*/IBS1* over EBS1*/dIBS1* is conveyed by the slower and more prominent $k_{off,2}$ of the RNA-RNA contact. The fast $k_{off,1}$ is likely to originate from an RNA conformation

where no Mg^{2+} is bound, while the slower $k_{off,2}$ links to a contact, which is stabilized by one or multiple site-bound Mg^{2+} ions. Since Mg^{2+} binding is transient, these states are exchanging on the timescale of imaging.

4.2.5 Degenerate FRET state is kinetically resolvable

To identify all interconversion rates between bound and unbound species, we used a maximum-likelihood approach that trains a global hidden Markov model (HMM) on all dynamic FRET traces.[182] We tested various kinetic networks and found a sequential three-state model with one zero and a twofold degenerate, high FRET state describing the data best whenever Mg^{2+} is present (Fig. 4.3b, Table 4.1, and Supplementary Fig. 4.16 and Supplementary Table 4.5, see also Kinetic model selection). The most appropriate model has been selected based on the Bayesian information criterion (BIC, Supplementary Fig. 4.17) and prior biochemical knowledge about the system as outlined below. Furthermore, it is critically evaluated by re-simulating the obtained rate system and by reproducing experimental dwell time distributions (Supplementary Figs. 4.18 and 4.19 and Supplementary Table 4.6). Biochemically, the kinetic model couples IBS1* or dIBS1* binding to Mg^{2+} coordination and can be interpreted in the following way: docking of the complement to the hairpin involves a detectable change in FRET efficiency, followed by a state transition, which is insensitive to FRET, thus degenerate, but kinetically resolvable. We attribute this second transition to one or multiple Mg^{2+} ions that specifically coordinate to the exon bound hairpin. Two key observations support this hypothesis: (i) in the absence of Mg^{2+}, the mechanism simplifies to a two-state system with one on- and one off-rate (Fig. 4.3a and Table 4.1); (ii) as seen from the NMR structure, a patch of negative electrostatic surface potential is exposed once IBS1* or dIBS1* binds to EBS1*, thereby forming a cavity where Mg^{2+} can coordinate. This pocket is particularly pronounced in the RNA-RNA interaction and absent when no exon is bound. There are two alternative metal ion-binding modes at the exon-intron interface between EBS1*/dIBS1* (Fig. 4.4a): an outer-sphere bound ion has been proposed based on nuclear Overhauser enhancement (NOE) restraints from the nucleic acids to the NMR-active $[Co(NH_3)_6]^{3+}$ (green sphere).[292] Alternatively, a Mg^{2+} may exchange one or more of its coordinated water molecules with RNA atoms and bind deeper within the pocket (dark green sphere). It seems likely that inner-sphere binding of Mg^{2+} in

that tunnel locks the tertiary contact in a rigid and stable conformation, which prevents fast dissociation of IBS1*. In the HMM model, the corresponding exchange rate $k_{\mathrm{ex},12}$, which describes the Mg^{2+}-induced stabilization of IBS1*, increases at the expense of the dissociation rate $k_{\mathrm{off},10}$ over the course of the Mg^{2+} titration (Supplementary Fig. 4.16). In the RNA-DNA hybrid, on the other hand, the off-rate outweighs the exchange between different EBS1*/dIBS1* conformations, thereby attenuating the degree of heterogeneity in dIBS1* unbinding.

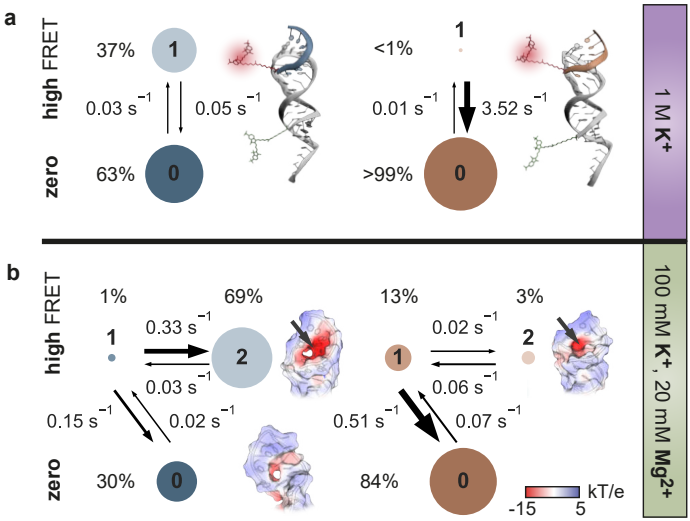


Fig. 4.3 | **Kinetic model of metal ion-induced IBS1* and dIBS1* recognition.** **a** Two-state binding of IBS1* (blue) and dIBS1* (orange) to the EBS1* hairpin under high monovalent ionic strength (0: exon unbound; 1: exon bound). The area of the circle is proportional to the state population (specified in percentage) and the arrow width reflects the relative rate of the transition. Binding rates, $k'_{\mathrm{on},01}$, are indicated as pseudo-first-order rates coefficients calculated from the trained HMM. **b** Three-state model with two degenerate high FRET states reflecting the coordination of Mg^{2+} to the formed tertiary contact (0: exon unbound, Mg^{2+} unbound; 1: exon bound, Mg^{2+} unbound; 2: exon bound, Mg^{2+} bound). Insets show surface representations of the RNA-RNA and RNA-DNA interaction, color coded by the electrostatic potential (red: −15 kTe; blue 5 kTe^{-1}). Arrows indicate coordination sites for Mg^{2+}.

4.2.6 Sugar puckering increases dynamics at the RNA-DNA interface

What is the molecular origin of the lower stability of the RNA-DNA contact and the reduction in kinetic heterogeneity? Structurally RNA and DNA are set apart by a methyl group on thymine and the lack of a 2′-OH on the DNA sugar. Differences in duplex stability have been attributed to both features.[300] The 2′-hydroxyl group contributes to the structural diversity of RNA by forming hydrogen bonds that make up tertiary structure elements such as the ribose zipper or the A-minor motif.[145,301] In helical regions, the puckering mode of the sugar is critically influenced by the presence or absence of the 2′-OH group. The most common helical topologies in nature are the A and B forms. While DNA mostly adopts the B-form (C2′-endo), RNA prefers the A-form (C3′-endo), in which the 2′-OH is in a sterically and electronically more favorable axial position (Fig. 4.4b).[302] The type of sugar pucker can be inferred from TOCSY experiments, where ^3J coupling between H1′ and H2′ protons is related to the dihedral angle via the Karplus equation.[302,303]. Coupling between the two nuclei is strong if the sugar is in a C2′-endo conformation with the protons in axial alignment. A switch to C3′-endo rotates the protons into an equatorial position where they are oriented perpendicularly to each other and coupling is weak.[291,292] Puckering dynamics have been observed on different timescales from picoseconds up to milliseconds.[304–306] The seven nucleotides of IBS1* and their base-paired complements on the EBS1* hairpin show no crosspeaks between H1′ and H2′ as opposed to the unpaired loop residues A10, U11, and U12, which display a sharp signal (Supplementary Fig. 4.20). In previous NMR structure calculations, the torsion angles have therefore been restrained to either a C3′-endo (IBS1* and paired nucleotides of EBS1*) or a C2′-endo (A10, U11, and U12) conformation.[291,292] In dIBS1*, correlations between H1′ and H2′/H2″ are observable for all seven residues, yet the linewidths of these peaks are broader compared to U11 and U12. Quantification of the peak volumes yields intermediary values, suggesting a fast exchange between puckering modes on the NMR timescale. Consequently, no restraints have been applied to these residues and the resulting lowest energy NMR structures show puckers from C2′-exo to C2′-endo (Fig. 4.4c). To assess whether the intermediary ring puckers (C4′-exo, O4′-endo, C1′-exo) are a result of interconverting conformations averaged out in the NMR measurement, we run molecu-

lar dynamics simulations for both tertiary contacts in their docked form. We first simulated the two contacts in the presence of K^+ only, and in a second set of runs, we added Mg^{2+} at the positions expected from NMR chemical shifts, NOE restraints, and electrostatic calculations (Supplementary Fig. 4.21). In both cases, the ribose of the IBS1* residues is predominantly in the canonical C3′-endo conformation as expected from the absence of IBS1* crosspeaks in the TOCSY spectrum (Fig. 4.4c and Supplementary Figs. 4.22–4.25). Excursions to a C2′-endo occur mostly at the flexible 3′-terminus (C65). Overall, the canonical A-form helix of EBS1*/IBS1* is relatively rigid.[302,305] In contrast, residues in dIBS1* undergo frequent repuckerings with correlation times in the mid picosecond range both in the presence and absence of Mg^{2+} (Supplementary Figs. 4.22 and 4.24). Their pucker profiles are very broad and cover most of the Eastern half of the pseudorotation cycle. The simulations show no significant energy barrier separating north and south puckers and thus a variety of intermediary deoxyribose conformations are explored (Supplementary Figs. 4.23 and 4.25). Interestingly, the unpaired loop residues A10-U12 of the EBS1* hairpin also switch between C2′-endo and C3′-endo conformations, but on a timescale of several nanoseconds (Supplementary Figs. 4.26 and 4.27). Unlike the residues in dIBS1*, intermediary puckers in A10-U12 are unfavored due to the presence of a distinct energy barrier. Because the dihedral angles of the sugar ring are all interdependent, the pucker dynamics propagate onto the backbone where they alter the distance between neighboring phosphates and the shape of the helix.[302] As a result, the RNA/DNA hybrid adopts neither a pure A-form nor a B-form geometry.[291,307,308] The conformational exchange at the RNA-DNA interface probably weakens the binding, thus contributing to the lower stability of hybrid contacts in general and in our case to faster dissociation rates of dIBS1*.[293,294]

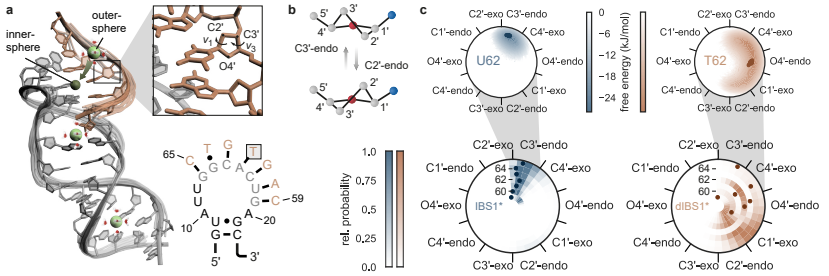


Fig. 4.4 | **Fig. 4 Sugar pucker distributions from all-atom MD simulations.** **a** NMR structure of EBS1*/dIBS1* (PDB: 2m1v[292]) with overlaid backbones of the 18 lowest energy structures. Outer-sphere binding sites based on NOE restraints from dIBS1* protons (i.e. H3 and CH_3 of T62/64 and H5/6 of C65) to $[Co(NH_3)_6]^{3+}$ (large green spheres; H_2O molecules are added for clarity) and a potential inner-sphere coordination mode where Mg^{2+} binds further inside the tunnel (smaller dark green sphere).[292] **b** Switch between energetically favored sugar pucker conformations C3′-endo and C2′-endo. **c** Circular histogram of pseudorotation angles along the seven nucleotides of IBS1* or dIBS1* (C59-C65) from MD simulations with 100 mM K^+ and 20 mM Mg^{2+}. Insets show distribution of pucker phases and amplitudes of U62 and T62. Puckers of the lowest energy NMR structures and their medians are indicated as dots for each residue.

4.3 Discussion

Group II introns have developed a highly specific recognition mechanism to cut and paste themselves from one part of the genome to another. Each of the three classes, IIA, IIB, and IIC, relies on a slightly different set of tertiary interactions to position the target strand in the active site for cleavage.[267,268] However, common to all is the EBS1, which forms a duplex directly adjacent to the 5′-splice site. Here, we have investigated the thermodynamics and kinetics governing the stability of this tertiary contact using single-molecule FRET in combination with molecular dynamic simulations.

Our FRET trajectories capture binding and unbinding events across various K^+ and Mg^{2+} concentrations. Consistent with previous experiments, IBS1 (un)docking and Mg^{2+} coordination are coupled, resulting in Mg^{2+} exchange rates on a timescale of milliseconds to seconds, thus accessible to camera-based detection.[175,280] We find that Mg^{2+} association and dissociation from the EBS1*/IBS1* contact are linked to very subtle conformational changes within the duplex, indiscernible by FRET. As a spectroscopic ruler, the FRET efficiency fails to resolve these kinetic states. In the time dimension, however, the states can be distinguished with regard to their dwell times (Fig. 4.5). Hence, exon unbinding is characterized by multiexponential decays from one degenerate FRET state, conceptually known as kinetic heterogeneity. Heterogeneity is common to many nucleic acid interactions and increases the complexity of the kinetic analysis16. Here, we have shown how to dissect such a state degeneracy and elucidate its molecular origin. For this purpose, we have drawn a state network of EBS1*/(d)IBS1* using a combination of dwell time analysis and hidden Markov modeling on an ensemble of FRET trajectories. Based on our model system, we propose that the 5′-exon dissociates at different rates depending on whether the Mg^{2+} binding sites are occupied or not. While outer-sphere Mg^{2+} ions are typically thought to exchange on a millisecond timescale[309], our previous single-molecule Mg^{2+} pulse experiments suggest that some ions remain associated with the RNA for seconds or even minutes.[280] These ions are probably chelated by the RNA[94] and buried in a binding pocket formed upon docking of the 5′-exon.[291] The said cavity is shallower in the hybrid EBS1*/dIBS1* structure and Mg^{2+} binding to RNA-DNA contacts is generally weaker.[292] Yet, Mg^{2+} is required to stabilize the intrinsically labile RNA-DNA contact. Loss of a coordinated Mg^{2+} ion is

therefore likely to be followed by immediate strand dissociation. Such a concerted mechanism of a Mg^{2+} and exon dissociation is consistent with the observed reduction in kinetic heterogeneity from IBS1* to dIBS1*. How does ribozyme catalysis benefit from heterogeneity? Group II introns have evolved to find just the right balance of keeping a correct substrate in the active site long enough for the two transesterifications to occur, while releasing mismatched sequences before the first chemical step can take place.[278] To achieve such high selectivity, ribozymes make use of kinetic partitioning. Mg^{2+} prolongates the residence time of the exon given that there is a suitable binding site. Selective stabilization of the structurally and catalytically important exon-intron tertiary interaction forms the basis of its heterogeneous kinetics. A depletion of Mg^{2+} homogenizes the rates. Furthermore, splicing is downregulated, because K^+ alone at physiological concentrations is neither able to stabilize the exon-intron contact sufficiently nor activate the phosphodiester bond for cleavage.[12,276]

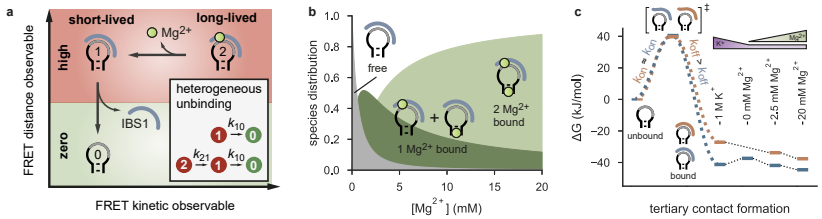


Fig. 4.5 | **Tertiary contact formation monitored by FRET in space and time.** **a** Exon release is described by two FRET observables: dye distance and state dwell times. Intron-exon dissociation can occur from state 1 (Mg^{2+} free) or state 2 (Mg^{2+} bound) and is thus kinetically heterogeneous. **b** Occupation of Mg^{2+} binding sites as a function of the ion concentration using Mg^{2+} binding constants from NMR chemical shift perturbations (Equation 4.6).[280] Cartoons depict possible species with no (gray), one (dark green), or two Mg^{2+} bound (light green). The Mg^{2+} free species corresponds to the high FRET state 1 in a, while the other species together make up state 2. Diffuse Mg^{2+} ions are omitted for clarity. **c** Free energy diagram of tertiary contact formation featuring a reactant-like, unstructured transition state. Off-rates determine the stability of the interactions. Dotted black lines connect levels of the bound state at different metal ion concentrations.

In the cell, where Mg^{2+} is present in very low millimolar concentrations, Ca^{2+} is tightly regulated, and other divalent metal ions are generally scarce, intron-encoded or host supplied proteins are recruited to stabilize the weak interactions and promote catalysis. The group IIB intron *Sc*.ai5γ from which the EBS1*/ IBS1* model interaction is derived, is a splicing-only intron, and resides in the housekeeping gene COX1 of yeast mitochondria, where it is assisted by splicing factors such as the ATP-dependent helicase Mss116.[135,310,311] Mobile group II introns, in contrast, still have their own open reading frame encoding a maturase that stabilizes the exon-intron contact.[27,312] Only recently, cryo-EM structures of a maturase-intron complex from Lactococcus lactis and Thermosynechococcus elongatus have revealed how the exon is sandwiched between the protein's thumb domain and EBS1/2.[289,290] This ternary complex between intron, exon, and protein might be of particular importance to convey stability to the labile RNA-DNA interaction during reverse splicing. Notably, the intron-encoded protein (IEP) has been found to bind more tightly to an intron lariat with the 5′-exon added in trans than to a spliced intron lacking the EBS/IBS interactions[312]. The exon-intron recognition complex and the IEP thus seem to cooperate to resolve instabilities in DNA target binding.

Our molecular simulations provide an atomic view on the exon-intron binding interaction and suggest that fast sugar puckering dynamics may be responsible for the lower affinity of the intron towards DNA exons. The simulations show that sugar puckers of dIBS1* are dispersed across the Eastern half of the pseudorotation cycle. In this energy basin the sugars are free to adopt different conformations. Does this switching have any regulatory function? On a microsecond to millisecond timescale, exchange between 3′-endo and 2′-endo sugars was found to be associated with helical to non-helical transitions, leading to alternative secondary structures, a feature common in riboswitches.[256,304] These rather slow puckering modes are further reminiscent of the cis/trans isomerization in peptide bonds containing proline residues, which were recently discovered to be a source of heterogeneity in intrinsically disordered protein interactions.[177] Unlike proline isomerization and secondary structure rearrangements, sugar puckering in dIBS1* happens on a timescale of pico to nanoseconds and is thus much faster than exon release. Consequently, the intron is prone to let go of its DNA substrate unless the contact is properly stabilized by Mg^{2+} and/or a protein. We see this facile reversibility of EBS1/dIBS1 binding as a sort of quality control

mechanism that ensures correct integration site selection as the ribonucleoprotein (RNP) scans along the exon.[313]
To test this hypothesis, the EBS1/IBS1 interaction needs to be put back into the context of the full intron. Such large and architecturally complex RNAs and RNPs have long eluded a detailed single-molecule kinetic analysis mainly because the site-specific introduction of suitable reporters, like fluorophores, has proven to be challenging.[314] For instance, to probe the same reaction coordinate as in our EBS1*/(d)IBS1* model construct, a dye needs to be placed adjacent to EBS1, which is located around nucleotide position 330 in *Sc.*ai5γ. Making such a FRET construct would require multiple fragments to be ligated enzymatically. Recent advances in co- and post-transcriptional nucleic acid labeling now provide more flexibility in choosing dye positions, which are close to functionally important elements, and thus provide a basis for studying the kinetics of large RNA-protein assemblies on a single-molecule level.[245,315]
Here, we have used instead a minimal construct, which is chemically synthesized and covalently labeled, giving us a more straightforward access to studying its binding kinetics. We have shown that ion binding and sugar puckering are two interacting processes that shape kinetic heterogeneity. While site specifically coordinated Mg^{2+} ions delay strand dissociation, sugar switching effectively promotes it. Off-rates are therefore the determining factor for contact stability, in line with previous studies on RNA kissing loops and various kinds of RNA, DNA, and hybrid duplexes.[295,316] A question that remains is whether sugar puckering attenuates heterogeneity by directly interfering with metal ion coordination. To explore the relative binding affinities of Mg^{2+} to pure RNA and RNA-DNA hybrid contacts, MD simulations with enhanced sampling may prove useful.[299,317] Such simulation might also detect small structural changes induced upon Mg^{2+} coordination that FRET struggles to resolve. Altering the degree of residual dynamics at binding interfaces is a common type regulation in biomolecules. By balancing flexibility and rigidity, sugar puckering and metal ions control RNA dynamics and form the molecular basis of kinetic heterogeneity.

Table 4.1 | Kinetic parameters determined with a hidden Markov model on dynamic single molecules.

Exon	**Metal ions**	$k'_{\mathrm{on},\mathbf{01}}$ ($10^{-2}\,\mathrm{s}^{-1}$) [a]	$k_{\mathrm{on},\mathbf{01}}$ ($10^{5}\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$) [b]	$k_{\mathrm{off},\mathbf{10}}$ (s^{-1})	$k_{\mathbf{12}}$ ($10^{-2}\,\mathrm{s}^{-1}$)	$k_{\mathbf{21}}$ ($10^{-2}\,\mathrm{s}^{-1}$)	K_{d} (nM) [c]	ΔG_{bind} ($\mathrm{kJ\,mol}^{-1}$)	$\Delta^{\ddagger}$ ($\mathrm{kJ\,mol}^{-1}$)
IBS1*	1 M K^{+} [d]	2.96(4)	8.45(11)	0.051(1)	–	–	59.0	-41.2	39.2(1)
IBS1*	20 mM Mg^{2+}	2.14(7)	6.11(20)	0.146(6)	33.1(1.4)	2.56(10)	15.2	-44.6	40.0(1)
dIBS1*	1 M K^{+} [d]	0.86(3)	1.72(6)	3.52(11)	–	–	17500	-27.2	43.1(1)
dIBS1*	20 mM Mg^{2+}	7.07(3)	14.1(1)	0.511(2)	2.35(7)	6.33(16)	255	-37.6	37.9(1)

Errors are standard deviations computed from likelihood ratio tests.[318]

[a] Pseudo-first-order association rate coefficients, $k'_{\mathrm{on},01}$, are extracted from the HMM

[b] Second-order rate coefficients, $k_{\mathrm{on},01}$, are calculated from the pseudo-first-order association rate coefficient and the total IBS1* concentration in solution (35 nM IBS1* or 50 nM dIBS1*). For IBS1* this is $k_{\mathrm{on},01} = k'_{\mathrm{on}} \times c^{-1}_{\mathrm{IBS1*}}$

[c] Dissociation constants are calculated from the relative state population S_{bound} and S_{unbound} and the ligand concentration. For IBS1* the K_{d} is thus given by $K_{\mathrm{d}} = S_{\mathrm{unbound}} \times c_{\mathrm{IBS1*}} \times S^{-1}_{\mathrm{bound}}$

[d] EDTA (100 µM) are added to chelate any traces of divalent metal ions

4.4 Methods

4.4.1 Construct design

Oligos were purchased from IBA Lifesciences (Göttingen, DE) or Microsynth (Balgach, Switzerland) with cyanine labels at the 5′ end (Cy3-EBS1* and Cy5-IBS1* or Cy5-dIBS1*) and a biotin at the 3′ end. Photophysical parameters (fluorescence lifetime, time-resolved anisotropy) of the labeled oligos have been previously characterized in detail.[171] Sequences are derived from the group IIB intron *Sc*.ai5γ and feature a slightly elongated stem as well as two A-to-C transversions in the loop. These were previously introduced at residues 15 and 17 to stabilize the exon-intron interaction (see Supplementary Methods). For surface immobilization on the microscopy slides, a single-stranded overhang was added that contains four uracils and a biotin at the 3′ terminus.[280] We refer to these mutated and labeled constructs as EBS1*/IBS1* (RNA-RNA) and EBS1*/dIBS1* (RNA-DNA) to differentiate them from the original sequence in the group II intron.

4.4.2 Single-molecule FRET experiments and trace processing

RNA hairpins and their RNA or DNA complements were buffered in 50 mM 3-(N-morpholino)propanesulfonic acid (MOPS) at pH 6.9. The imaging solution further contained 100 mM KCl, 1% glucose (w/v), 1 mM Trolox, an oxygen-scavenging system (2170 U mL^{-1} catalase, 165 U mL^{-1} glucose oxidase), and varying amounts of $MgCl_2$ (0 mM–20 mM) or alternatively 1 M KCl with 100 µM EDTA to remove any traces of divalent ions. The biotinylated EBS1* RNA was immobilized on a quartz surface via a streptavidin linkage.[261] For this purpose, the microfluidic channel was first coated with biotinylated bovine serum albumin (5 min), treated with streptavidin (10 min), and finally incubated with biotinylated EBS1* RNA (10 pM, 5 min). IBS1* or dIBS1* strands were then flushed into the chamber at concentrations of 35 or 50 nM, respectively. Molecules were imaged on a custom-built TIRF microscope using a water-immersion objective (UPlanSApo 60×/1.2-W, Olympus) at a camera frame rate of 5 or 10 Hz (Andor iXon DU-897). Single molecules with anti-correlated donor/acceptor emission were selected and corrected for background and donor bleedthrough into the acceptor channel using the freely available software package MASH-FRET (https://github.com/RNA-FRETools/MASH-FRET.git).[175,319]

The FRET efficiency was then calculated from the corrected donor and acceptor intensities after donor excitation according to

$$FRET = \frac{I_{\mathrm{A}}}{I_{\mathrm{A}} + I_{\mathrm{D}}} \tag{4.1}$$

FRET histograms were fitted to a Gaussian mixture model and uncertainties were estimated by bootstrapping.[320]

4.4.3 Single-molecule kinetic analysis

Individual single-molecule traces were discretized based on a threshold criterion to generate sets of dwell times in the well-separated zero and high FRET state. A mean dwell time, $\langle t_{\mathrm{zero},n} \rangle$ and $\langle t_{\mathrm{high},n} \rangle$ was calculated for each trace n. On the other hand, cumulative dwell time histograms were computed by sorting dwell times from all dynamic molecules. Only molecules showing at least two transitions between the two states were considered as dynamic. Characteristic decay constants, $\tau_{\mathrm{zero},m}$ and $\tau_{\mathrm{high},m}$, were obtained by fitting a biexponential (Supplementary Equation 4.8) or a stretched exponential (Supplementary Equation 4.9) to the distribution. For the binding reaction, second-order rate coefficents, kon, were calculated by dividing the pseudo-first-order on-rate, k'_{on}, by the total IBS1* or dIBS1* concentration in solution: $k_{\mathrm{on}} = k'_{\mathrm{on}} \times c^{-1}_{(\mathrm{d})\mathrm{IBS1}*}$. Dissociation constants, K_{d}, were computed from the on and off-rates for each decay component m or alternatively for each individual molecule n. In case of the latter, cumulative distributions of $K_{\mathrm{d},n}$ were built and fitted to a normalized logistic function

$$P(K_{\mathrm{d},n}) = \frac{1}{1 + (K_{\mathrm{d}}/K_{\mathrm{d},n})^p} \tag{4.2}$$

Error estimates on all variables were computed from a set of 100 bootstrap samples. For a more detailed mathematical description of the kinetic analysis, we refer the reader to the Supplementary Information. The Gibbs free energy of the bound state with respect to the free hairpin was determined using

$$\Delta G = RT \ln(K_{\mathrm{d}}) \tag{4.3}$$

with the gas constant R and temperature T. The energy barrier separating the bound and unbound state was computed according to transition state theory as

$$\Delta G^{\ddagger} = -RT \ln\left(\frac{k_{\mathrm{on}} h}{\kappa k_B T}\right) \tag{4.4}$$

with the binding rate k_{on}, Planck's constant h, and Boltzmann's constant k_B. Our calculations of $\Delta G^{\ddagger}$ assume a transmission coefficient $\kappa = 1$, thus providing a lower limit for the difference in free energy between the reactants and the transition state.[226,321] The fraction of Mg^{2+}-bound RNA Θ is computed using association constants, K_1 and K_2, derived from NMR chemical shift perturbations[280]

$$\Theta = \frac{K_1[\mathrm{Mg}^{2+}] + K_2[\mathrm{Mg}^{2+}] + K_1K_2[\mathrm{Mg}^{2+}]^2}{1 + K_1[\mathrm{Mg}^{2+}] + K_2[\mathrm{Mg}^{2+}] + K_1K_2[\mathrm{Mg}^{2+}]^2}. \tag{4.5}$$

4.4.4 Global HMM and FRET trace simulation

For each salt condition, a HMM was first trained on a trace-by-trace basis as implemented in the software package SMACKS.[182,318] In a second step, the predetermined emission probabilities were fixed and the start and transition probabilities of a global HMM were optimized collectively on the entire ensemble of dynamic FRET traces. Kinetic models with different connectivities and degeneracies were tested and evaluated based on a BIC and biochemical significance (see Kinetic model selection). To validate the obtained rate system, degenerate FRET traces were re-simulated in a subroutine of MASH-FRET[322] from the underlying rate matrix

$$\mathbf{K} = \begin{pmatrix} -k_{01} & k_{01} & 0 \\ k_{10} & -(k_{10} + k_{12}) & k_{12} \\ 0 & k_{12} & -k_{12} \end{pmatrix} \tag{4.6}$$

Dwell time histograms were built analogously to the experimental traces.

4.4.5 Poisson-Boltzmann continuum electrostatics

The electrostatic surface potential of EBS1* alone and EBS1* in complex with either IBS1* or dIBS1* were computed with the APBS Electrostatics plugin[323] for PyMOL. Partial charges and VdW radii were assigned with PDB2PQR[324] using parameters from AMBER-ff99. Calculations were run at 300 K with 100 mM monovalent and 20 mM divalent ions.

4.4.6 Molecular dynamics simulation

NMR structures of the tertiary contact formed between EBS1* and IBS1* or dIBS1* (PDB: 2M23 and 2M1V) were used as starting points

for molecular dynamics simulations with GROMACS 5.1.[325] The simulations use the AMBER-ff99 force field[326] with parmbsc0[327] and χOL3[328,329] corrections (corresponds to AMBER-ff14). Monovalent ions were modeled with parameters from Joung and Cheatham[330], while those for Mg^{2+} were taken from Li et al.[331] The molecules were solvated in a dodecahedral box filled with TIP3P water and randomly interspersed with K^+ and Cl^- ions to neutralize the net charge and reach a concentration of about 100 mM. In a second set of simulations, Mg^{2+} ions were placed at specific locations suggested by NMR chemical shifts and NOE restraints to $[Co(NH_3)_6]^{3+}$. The nucleic acids were then charge neutralized and the K^+ and Mg^{2+} concentrations set to 100 mM and 20 mM, respectively (concentrations are estimates because of the limited box size). In an equilibration phase (2×5 ns, NVT and the NPT ensembles), the temperature was adjusted to 300 K using the velocity rescale thermostat and the pressure was kept constant at 1 bar by the Parrinello-Rahman barostat. Bonds were constrained by the LINCS algorithm with an integration time step of 2 fs. Nonbonded interactions use the Verlet scheme with a cut-off of 1.4 nm, while long-range electrostatics were treated with the particle mesh Ewald algorithm. Simulations of EBS1*/IBS1* and EBS1*/dIBS1* (both with and without Mg^{2+}) were run for 2 µs each. Convergence was assessed by computing the root mean square deviation over the nucleic acids and time binned histograms of the sugar puckers, showing that the repuckering time is much shorter than the overall simulation length. The pseudorotation angles of the riboses were calculated with PLUMED 2.5.2[332] and assigned to one of ten puckering modes (phase angle increments of 36°)[302] according to the definition by Huang et al.[333] (Supplementary Equations 4.21-4.23), which are related to the one introduced originally by Altona and Sundaralingam.[334] The spatial distribution of Mg^{2+} ions, starting from a set of experimentally observed positions, is visualized using a Gaussian kernel density estimate. The point cloud is normalized by the ion concentration, color coded by the density, and mapped onto the NMR structure (Supplementary 4.21).

4.4.7 NMR linewidth analysis

Linewidths and peak volumes in [^{1}H-^{1}H]-TOCSY spectra were calculated using NMRFAM-Sparky[335] by fitting a set of two-dimensional Gaussians to the assigned peaks.

4.5 Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request. The source data underlying Figs. 4.1d–g, 4.2a–g, 4.3a/b, 4.4c, 4.5b/c and Supplementary Figs. 4.6, 4.7, 4.12, 4.14, 4.20, 4.22–4.27 are provided as a Source Data file.

4.6 Code availability

A Matlab-based software package for video processing, evaluation of single-molecule trajectories, histogram, and dwell time distribution analysis is freely available at https://github.com/RNA-FRETools/MASH-FRET.git. Documentation of the entire workflow of MASH-FRET can be found at https://rna-fretools.github.io/MASH-FRET/.

4.7 Acknowledgments

We thank Ben Schuler, Helmut Grubmüller and Mauro Schilling for discussions on single-molecule kinetics and molecular dynamics simulations; Silke Johannsen for suggestions on NMR data analysis; Sonja Schmid for inputs on the ensemble HMM; Christian G. Hübner for providing computational resources for data analysis. This work was supported by the European Research Council (ERC Starting Grant MIRNA No. 259092, to R.K.O.S.) and the Forschungskredit of the University of Zurich (FK-17-098 to F.D.S., FK-14-096/15-96 to R.B.). We are thankful for further financial support by the Swiss National Science Foundation (to R.K.O.S.) and the Stiftung für Wissenschaftliche Forschung (to R.K.O.S. and R.B.) of the University of Zurich.

4.8 Author contributions

F.D.S., D.K., R.B., and R.K.O.S. designed the research. M.K. collected single-molecule FRET data. F.D.S., M.K., and R.B. analyzed and interpreted the data. F.D.S. and R.A.C. ran and analyzed MD simulations. R.B., D.K., and R.K.O.S. supervised the work. F.D.S. and R.B. wrote the paper with the help of all authors.

4.9 Supplementary Methods

4.9.1 Construct design

The exon and intron binding site 1 (EBS1/IBS1) sequences used herein are derived from the group IIB intron *Sc*.ai5γ. Compared to the wild type, the stem was elongated by four nucleotides and a 3′- overhang was introduced for surface immobilization (underlined residues). Furthermore, two point mutations in the loop were introduced to stabilize the tertiary interaction in vitro (bold residues) without affecting the cleavage rate.[336] A star is appended to the name in order to differentiate the constructs from their native counterpart in the group IIB intron.

EBS1* Cy3–5′-GGAGUAUGUAUUGGCACUGAGCAUACUCCUUUU-3′–biotin

IBS1* Cy5 – 5′-CAGUGUC-3′

dIBS1* Cy5 – 5′-CAGTGTC-3′

4.9.2 Single-molecule dwell time analysis

In the following we outline the procedure used to extract dwell times in the zero and high FRET state, their corresponding binding and unbinding rates as well as distributions of dissociation constants. The nomenclature used herein follows in most parts the one introduced by Kowerko et al., PNAS (2015).[280]

Thermodynamic parameters such as dissociation constants, K_d, or binding free energies ΔG can be calculated directly from the relative occupation of the two FRET states in the FRET histogram and the total ligand concentration (Supplementary Fig. 4.9a). Below, we decided to use a shot noise free approach to get both kinetic and thermodynamic parameters from the state dwell times. For this purpose, single-molecule time traces were discretized using a thresholding criterion corresponding to the midpoint between the centers of the zero and high FRET distribution. Complementary cumulative dwell time distributions $C(t_\mathrm{zero})$ and $C(t_\mathrm{high})$ were computed by first sorting all dwell times of all molecules recorded under equal buffer conditions in ascending order and then calculating the relative occurrence of dwell time $i \in \{1, ..., i_{\mathrm{max},n}\}$ of molecule $n \in \{1, ..., n_\mathrm{max}\}$. This gives a set of probabilities $p(t) = \{p_1, ...p_N\}$ for the unique dwell times $t = \{t_1 ...t_N\}$ from which the complementary

cumulative dwell time probability is computed as

$$C(t_j) = 1 - \sum_{l=1}^{j} p(t_l), \quad \text{for } j = 1, 2, ...N \tag{4.7}$$

where t_j is the jth element in the dwell time sets t_{zero} or $t$high.
The dwell time distribution $C(t_{\text{zero}})$ and $C(t_{\text{high}})$ can be described as a series of exponentials, i.e. for the high FRET state

$$C(t_{\text{high}}) = \sum_{m=1}^{m_{\max}} a_{\text{high},m}\, e^{t_{\text{high}}/\tau_{\text{high},m}} \quad \text{with} \sum_{m=1}^{m_{\max}} a_{\text{high},m} = 1 \tag{4.8}$$

Alternatively, a single-exponential is modified by a stretching factor β

$$C(t_{\text{zero}}) = e^{(t_{\text{zero}}/\tau_{\text{zero}})^{\beta}} \tag{4.9}$$

Binding k_{on} and unbinding rates k_{off} were calculated from time constants t_{zero} and t_{high} for each decay component m according to

$$k_{\text{on},m} = \frac{1}{\tau_{\text{zero},m}\, c_{\text{total(IBS1*)}}} \tag{4.10}$$

$$k_{\text{off},m} = \frac{1}{\tau_{\text{high},m}} \tag{4.11}$$

Dissociation constants of the respective decay components $K_{\text{d},m}$ were calculated from the on- and off-rates as follows

$$K_{\text{d},m} = \frac{k_{\text{off},m}}{k_{\text{on},m}} = \frac{\tau_{\text{zero},m}\, c_{\text{total(IBS1*)}}}{\tau_{\text{high},m}} \tag{4.12}$$

Dissociation constants may also be calculated for an individual molecule n by averaging over $i_{\max}$ and $j_{\max}$ dwell times in the zero and high FRET state respectively

$$K_{\text{d},n} = \frac{\langle t_{\text{zero},n} \rangle\, c_{\text{total(IBS1*)}}}{\langle t_{\text{high},n} \rangle} \tag{4.13}$$

$$\text{with } \langle t_{\text{zero},n} \rangle = \frac{\sum_{i=1}^{i_{\max}} t_{\text{zero},i,n}}{i_{\max,n}} \quad \text{and } \langle t_{\text{high},n} \rangle = \frac{\sum_{j=1}^{j_{\max}} t_{\text{high},j,n}}{t_{\max,n}} \tag{4.14}$$

The number of dwell times $i_{\max}$ and $j_{\max}$ can vary between different molecules and $|i_{\max} - j_{\max}| = 0$ or 1. In the limiting case where $i_{\max,n} = j_{\max,n} = 1$ the average dwell times $\langle t_{\text{zero},n}\rangle$ and $\langle t_{\text{high},n}\rangle$ would be exponentially distributed according to

$$p(\langle t_{\text{zero},n}\rangle) = e^{\langle -t_{\text{zero},n}\rangle/\tau_{zero}} \quad \text{and } p(\langle t_{\text{high},n}\rangle) = e^{\langle -t_{\text{high},n}\rangle/\tau_{high}} \tag{4.15}$$

yet for most molecules $\langle t_{\text{zero},n}\rangle$ and $\langle t_{\text{high},n}\rangle$ are averaged over $i_{\max,n}$ and $j_{\max,n} > 1$ dwell times which is described by a gamma distribution[337]

$$p_\Gamma(\langle t_{\text{zero},i_{\max,n}}\rangle) = A\frac{k^{i_{max}}\,\langle t_{\text{zero},i_{\max,n}}\rangle^{i_{max}-1}}{\Gamma(i_{\max})}e^{-k\langle t_{\text{zero},i_{\max,n}}\rangle} \tag{4.16}$$

with amplitude A and rate $k = \tau_{\text{zero}}^{-1}$. Γ is the gamma function and $i_{\max,n}$ the number of dwell times over which is being averaged. The higher the number of dwell times per trace, the narrower the distribution will get. Hence, the distribution of $K_{\text{d},n}$ values can be expressed as a ratio of two gamma distributions

$$\begin{aligned} p(K_{\text{d},n}) &= \frac{p_\Gamma(\langle t_{\text{zero},i_{\max,n}}\rangle)\, c_{\text{total(IBS1*)}}}{p_\Gamma(\langle t_{\text{high},j_{\max,n}}\rangle)} \\ &= A\frac{K_{\text{d}}^{-N}}{B(N,N)}\big(1+\frac{K_{\text{d},n}}{K_{\text{d}}}\big)^{-2N}K_{\text{d},n}^{N-1} \end{aligned} \tag{4.17}$$

with $N := i_{\max} = j_{\max}$, as described previously by Coelho and Mexia.[338] Here, $B(i_{\max}, j_{\max})$ refers to the beta function and A is an amplitude to account for the frequency of the $K_{\text{d},n}$ values.
Expressing supplementary equation 4.17 in terms of $p(\log K_{\text{d},n})$ gives

$$\begin{aligned} p(\log K_{\text{d},n}) = A\frac{(10^{\log K_{\text{d}}})^{-N}}{B(N,N)}\Big(1+\frac{10^{\log K_{\text{d},n}}}{10^{\log K_{\text{d}}}}\Big)^{-2N} \\ (10^{\log K_{\text{d},n}})^{N-1}\,10^{\log K_{\text{d},n}}\ln(10) \end{aligned} \tag{4.18}$$

In order to make the distribution independent of the bin size, cumulative distributions $P(\log K_{\text{d,n}})$ were built analogously to the cumulative dwell time distribution and fitted to a normalized logistic function

$$P(K_{\text{d},n}) = \frac{1}{1+(K_{\text{d}}/K_{\text{d},n})^p} \tag{4.19}$$

$$P(\log K_{\text{d},n}) = \frac{(10^{\log K_{\text{d},n}-\log K_{\text{d}}})^p}{1+(10^{\log K_{\text{d},n}-\log K_{\text{d}}})^p} \tag{4.20}$$

The parameter, p describes the steepness of the cumulative distribution and is thus related to the number of dwell times N in supplementary equation 4.17/4.18.

4.9.3 Sugar pucker pseudorotation cycle

Pseudorotation angle and pucker amplitude were calculated according to Huang and York[333] and rely on two endocyclic torsion angles ν_1 and ν_3. The Cartesian coordinates

$$Z_x = \frac{\nu_1 + \nu_3}{2\cos(4\pi/5)} \tag{4.21}$$

$$Z_y = \frac{\nu_1 - \nu_3}{2\sin(4\pi/5)} \tag{4.22}$$

can be transformed into polar coordinates to give the pucker phase P and amplitude A

$$P = \arctan\left(\frac{Z_y}{Z_x}\right) \tag{4.23}$$

$$A = \sqrt{Z_x^2 + Z_y^2} \tag{4.24}$$

4.9.4 Rate matrix for FRET trace simulations

In a Markov chain, state transitions are treated as homogenous processes, and the respective $J \times J$ matrix $\mathbf{P}$ with transition probabilities p_{ij} is defined by the transition rates $k_{i\neq j}$ and the camera frame rate f as follows[322]

$$\mathbf{P} = \begin{pmatrix} e^{-\sum_{j=2}^{J} \frac{k_{1j}}{f}} & (1-p_{11})\frac{k_{12}}{\sum_{j=2}^{J} k_{1j}} & \cdots & (1-p_{11})\frac{k_{1j}}{\sum_{j=2}^{J} k_{1j}} \\ (1-p_{22})\frac{k_{21}}{\sum_{j\neq 2}^{J} k_{2j}} & e^{-\sum_{j\neq 2}^{J} \frac{k_{2j}}{f}} & & \\ \vdots & & \ddots & \vdots \\ (1-p_{JJ})\frac{k_{i1}}{\sum_{i\neq j}^{J} k_{ij}} & & \cdots & e^{-\sum_{i\neq j}^{J} \frac{k_{ij}}{f}} \end{pmatrix} \tag{4.25}$$

where $p_{i=j}$ is the probability to stay within the same state and $p_{i\neq j}$ the probability to transit from state i to state j, respectively. Rates are assumed to adopt values ≤ 0, thus the probabilities are within $1 \leq p_{i,j} \leq 0$. With the integration time of the camera $\Delta t = f^{-1}$, the first order Taylor-series approximation $T_1 p_{i=j}(\Delta t; 0) = p(0) + p'(0)\Delta t = 1 - \sum_{i\neq j}^{J} k_{ij}\Delta t$ yields the commonly used transition probability matrix for $k < f$

$$\mathbf{P} = \begin{pmatrix} 1-\sum_{j=2}^{J} k_{1j}\Delta t & k_{12}\Delta t & \cdots & k_{1j}\Delta t \\ k_{21}\Delta t & 1-\sum_{j\neq 2}^{J} k_{2j}\Delta t & & \\ \vdots & & \ddots & \vdots \\ k_{i1}\Delta t & & \cdots & 1-\sum_{i\neq j}^{J} k_{ij}\Delta t \end{pmatrix}. \tag{4.26}$$

The corresponding rate matrix is

$$\mathbf{P} = \begin{pmatrix} \sum_{j=2}^{J} k_{1j} & k_{12} & \cdots & k_{1j} \\ k_{21} & -\sum_{j\neq 2}^{J} k_{2j} & & \\ \vdots & & \ddots & \vdots \\ k_{i1} & & \cdots & -\sum_{i\neq j}^{J} k_{ij} \end{pmatrix}. \tag{4.27}$$

4.9.5 Coupling of tertiary contact formation and Mg^{2+} binding and selection of an appropriate kinetic model

Complex formation between a receptor (EBS1*) and a ligand (IBS1* or dIBS1*) is linked to Mg^{2+} binding in a thermodynamic cycle (Supplementary Fig. 4.9b).[296] It consists of four states, where the receptor exists in either unbound (U) or native (N) form in the absence or presence of Mg^{2+}. Historically, the coupling of Mg^{2+} binding and folding has often been described by the following chemical equilibrium

$$\mathrm{U} + m\,\mathrm{Mg}^{2+} \rightleftharpoons \mathrm{N} \bullet (\mathrm{Mg}^{2+})_m$$

where m refers to the number of discrete Mg^{2+} ions associated with the native state. The associated equilibrium constant is given by

$$K_{\mathrm{eq}} = \frac{[\mathrm{N} \bullet (\mathrm{Mg}^{2+})_m]}{[\mathrm{U}] + [\mathrm{Mg}^{2+}]^m} \tag{4.28}$$

which translates into the well-known Hill equation

$$\theta = \frac{[\mathrm{Mg}^{2+}]^m}{K_{\mathrm{eq}} + [\mathrm{Mg}^{2+}]^m} \tag{4.29}$$

where θ is the fraction of native EBS1*/IBS1* with bound Mg^{2+}. The limitations of the Hill model in describing Mg^{2+} assisted RNA folding have been widely discussed in the literature.[92,93,296,339] The model presumes that the native state has m more Mg^{2+} ions associated than the unbound one. However, charge neutrality is not satisfied by just a few site-bound Mg^{2+}. Instead, the overwhelming majority of Mg^{2+} are delocalized and remain hydrated in the ion atmosphere. Poisson-Boltzmann theory has proven useful to describe the electrostatic interactions of the ion atmosphere with a charged polyelectrolyte, like RNA or DNA. The software package APBS Electrostatics11 implements Poisson-Boltzmann theory and was used to compute the electrostatic surface potential of the EBS1* hairpin alone and in complex with IBS1* or dIBS1* (Fig. 4.9b). A patch of negative surface potential, representing a possible Mg^{2+} interaction site, is only exposed when IBS1* or dIBS1* bind to the hairpin (state N in the thermodynamic cycle, Supplementary Fig. 4.9b). The lack of such a pocket in the unbound hairpin makes a specific coordination of Mg^{2+} at those residues unlikely in the ligand free state (U). We thus excluded the $\mathrm{U} \bullet (\mathrm{Mg}^{2+})_m$ in the thermodynamic cycle from

our kinetic model. Thus, the transition from the free EBS1* hairpin to a stable tertiary contact involves first the binding of IBS1* or dIBS1* and secondly the coordination of a Mg^{2+} ion. Conversely, the reverse reaction implies release of the ion followed by undocking of the complement. In EBS1*/dIBS1* these two processes are likely concerted. Such biochemical knowledge about the system under study helps to find an appropriate kinetic model and can corroborate a purely statistical criterion such as the Bayesian information criterion (BIC, Supplementary Fig. 4.17). Originally, we tested different kinetic networks: two-state (0-1), cyclic three state (-0-0-1-, -0-1-1-), linear three-state (0-1-1, the one selected) as well as a cyclic four-state model: (-0-0-1-1-). In this syntax, the value 0 corresponds to the zero FRET state and the value 1 to the high FRET state. Degeneracy is implied if the same digit occurs more than once. Note that in the linear three state model the degenerate states are a priori not identical because they differ in the number of connections (Supplementary Fig. 4.17). After training the HMM, the linear three-state model was validated again by re-simulated degenerate FRET traces from the optimized rate system using MASH-FRET (see Supplementary Methods).[322] Dwell time histograms were built in the same way as in the experiment and then compared to the experimental ones (Supplementary Fig. 4.17 and Supplementary Fig. 4.19). Overall, the kinetic rates, k'_{on} and $k_{\mathrm{off},1/2}$, are in very good agreement between simulated and experimental decays, thus the HMM is able to disentangle the kinetic heterogeneity induced by divalent metal ions (Supplementary Table 4.6).[296] The contribution of magnesium binding to contact formation, $\Delta\Delta G_{\mathrm{Mg^{2+}}}$, can be mathematically expressed as

$$\Delta\Delta G_{\mathrm{Mg^{2+}}} = \Delta G^{\mathrm{N}}_{\mathrm{Mg^{2+}}} - \Delta G^{\mathrm{U}}_{\mathrm{Mg^{2+}}} = (\Delta G^{\mathrm{N}}_{\mathrm{D}} + \Delta G^{\mathrm{N}}_{\mathrm{S}}) - (\Delta G^{\mathrm{U}}_{\mathrm{D}} + \Delta G^{\mathrm{U}}_{\mathrm{S}}) \tag{4.30}$$

where the subscript D and S refer to the diffuse and site binding free energy terms to either the native (N) or unfolded (U) state.[296] Since site-binding to the loop region of the free hairpin (U) was not observed by NMR chemical shifts[336] and a discrete binding pocket is absent (see above), the term $\Delta G^{\mathrm{U}}_{\mathrm{S}}$ is omitted and $\Delta\Delta G_{\mathrm{Mg^{2+}}}$ reduces to

$$\Delta\Delta G_{\mathrm{Mg^{2+}}} = \Delta\Delta G_{\mathrm{D}} + \Delta G^{\mathrm{N}}_{\mathrm{S}} \tag{4.31}$$

where $\Delta\Delta G_{\mathrm{D}}$ is the difference in diffuse binding to the N and U state.

4.9.6 Analytical solution of a 1:1 ligand-receptor interaction

Binding of a ligand L to a single receptor site R can be described by the degree of receptor saturation θ as

$$\theta = \frac{[\mathrm{L}]}{K_\mathrm{d} + [\mathrm{L}]} = \frac{[\mathrm{L_{tot}}] - [\mathrm{RL}]}{K_\mathrm{d} + [\mathrm{L_{tot}}] - [\mathrm{RL}]} \tag{4.32}$$

with

$$[\mathrm{RL}] = \frac{K_\mathrm{d} + [\mathrm{L_{tot}}] + [\mathrm{R_{tot}}] \pm \sqrt{(K_\mathrm{d} + [\mathrm{L_{tot}}] + [\mathrm{R_{tot}}])^2 - 4[\mathrm{L_{tot}}][\mathrm{R_{tot}}]}}{2} \tag{4.33}$$

and $[\mathrm{RL}] > 0$.

Hence, the receptor saturation can be expressed in terms of the total ligand concentration $[\mathrm{L_{tot}}]$, total receptor concentration $[\mathrm{R_{tot}}]$, and the dissociation constant Kd. Binding isotherms for different combinations of these three parameters, depending on the experiment and system under study are displayed in Supplementary Fig. 4.8. The curves illustrate common restraints imposed by the experimental setup. In the smFRET experiments the amount of fluorescently labeled and freely diffusing ligand that can be added is limited because of background issues in the acceptor channel (direct excitation of Cy5 by the green laser is on the order of 5%). Consequently, the fraction of receptor saturation is low in the case of EBS1*/dIBS1*. In the NMR experiment, on the other hand millimolar amounts of receptor and ligand are required to achieve good signal to noise ratios. These concentrations are much higher than the expected K_d of the interaction $[\mathrm{L_{tot}}] \approx [\mathrm{R_{tot}}] \gg K_\mathrm{d}$. The resulting binding curve shows a sharp kink where $\mathrm{R_{tot}} = \mathrm{L_{tot}}$ and thus a small excess of ligand is sufficient to quantitatively saturate the receptor.

Table 4.2 | Kinetic parameters from mean dwell time analysis of single-molecule FRET experiments.

Cognate strand	**Metal ion concentration**	center of $\langle \boldsymbol{k}_{\mathrm{on}} \rangle$ distr. ($10^5\,\mathrm{M^{-1}\,s^{-1}}$) [a]	center of $\langle \boldsymbol{k}_{\mathrm{off}} \rangle$ distr. ($\mathrm{s^{-1}}$)	$\boldsymbol{K}_{\mathrm{d}}$ (nM) [b]	$\boldsymbol{\Delta G}_{\mathrm{bind}}$ ($\mathrm{kJ\,mol^{-1}}$) [b]	$\langle \boldsymbol{\Delta G}^{\ddagger} \rangle$ ($\mathrm{kJ\,mol^{-1}}$) [c]
IBS1*	1 M K^+ [d]	8.8	0.052	59 ± 3	-41.3 ± 0.1	39.1
IBS1*	0 mM Mg^{2+}	3.4	0.047	285 ± 113	-37.3 ± 1.0	41.4
IBS1*	0.5 mM Mg^{2+}	4.5	0.031	83 ± 22	-40.4 ± 0.7	40.7
IBS1*	2.5 mM Mg^{2+}	4.1	0.025	117 ± 23	-39.6 ± 0.5	41.0
IBS1*	5 mM Mg^{2+}	4.9	0.021	89 ± 19	-40.2 ± 0.5	40.5
IBS1*	10 mM Mg^{2+}	7.4	0.021	52 ± 12	-41.4 ± 0.6	39.5
IBS1*	15 mM Mg^{2+}	5.6	0.017	42 ± 11	-42.2 ± 0.6	40.2
IBS1*	20 mM Mg^{2+}	7.6	0.017	29 ± 8	-43.0 ± 0.7	39.4
dIBS1*	1 M K^+ [d]	2.3	2.6	8530 ± 1930	-28.9 ± 0.1	42.4
dIBS1*	2.5 mM Mg^{2+}	3.6	0.67	1630 ± 130	-33.0 ± 0.2	41.3
dIBS1*	5 mM Mg^{2+}	4.8	0.42	1110 ± 40	-34.0 ± 0.1	40.6
dIBS1*	10 mM Mg^{2+}	7.8	0.32	589 ± 18	-35.5 ± 0.1	39.4
dIBS1*	15 mM Mg^{2+}	9.2	0.33	447 ± 17	-36.2 ± 0.1	39.0
dIBS1*	20 mM Mg^{2+}	12	0.32	353 ± 8	-36.8 ± 0.1	38.3

Errors are standard deviations computed from likelihood ratio tests.[318] The center of the point cloud $\langle k_{\mathrm{on}} \rangle$ versus $\langle k_t extnormal of f \rangle$ is determined by k-means clustering.
All experiments were conducted with a monovalent ionic background of 100 mM ceKCl unless stated otherwise.
[a] Second-order mean rate coefficients, $\langle k_{\mathrm{on}} \rangle$, are calculated from the pseudo-first-order mean association rate coefficient and the total IBS1* concentration in solution (35 nM IBS1* or 50 nM dIBS1*). For IBS1* this is $\langle k_{\mathrm{on}} \rangle = \langle k'_{\mathrm{on}} \rangle \times c^{-1}_{\mathrm{IBS1*}}$. The tabulated values refer to the center of the point cloud of $\langle k_{\mathrm{on}} \rangle$ versus $\langle k_{\mathrm{off}} \rangle$.
[a] Dissociation constants, K_{d}, are calculated from the logistic fits (Eqn. S13) to the cumulative distribution of $K_{\mathrm{d},n}$ values. Binding free energies, ΔG_{bind}, are computed from the K_{d}.
[c] The free energy difference between the unbound and the transition state is calculated using the center of the distribution of mean association rates $\langle k_{\mathrm{on}} \rangle$.

Table 4.3 | Comparison of single-molecule kinetics (sm, from mean dwell times, Supplementary Table 4.2) with surface plasmon resonance (SPR) experiments (at 25 °C).[292]

Cognate strand	**Metal ion concentration**	center of $\langle k_{on} \rangle$ distr. ($10^5\,M^{-1}\,s^{-1}$) [a]	k_{on} ($10^5\,M^{-1}\,s^{-1}$) [a]	center of $\langle k_{off} \rangle$ distr. (s^{-1})	$\langle k_{off} \rangle$ (s^{-1}, sm)	K_d (nM, nm) [c]	K_d (nM, SPR) [b]
IBS1*	0 mM Mg^{2+}	3.4	1.0 ± 0.2	0.047	0.015 ± 0.010	285 ± 113	150 ± 30
IBS1*	5 mM Mg^{2+}	4.9	2.1 ± 1.3	0.021	0.006 ± 0.001	89 ± 19	30 ± 10
dIBS1*	0 mM Mg^{2+}	–	0.5 ± 0.2	–	1.45 ± 0.35	–	$29\,000 \pm 5700$
dIBS1*	5 mM Mg^{2+}	4.8	2.4 [c]	0.42	0.18 [c]	1110 ± 40	720 [c]

Errors of single-molecule experiments (if specified) are standard deviations of 100 bootstrap samples.[320] Errors of SPR experiments (if specified) are standard deviations of independent measurements on two sensor chips. Experiments were conducted with a monovalent ionic background of 100 mM KCl.

[a] Second-order rate coefficients, $k_{on,01}$, are calculated from the pseudo-first-order association rate coefficient and the total IBS1* concentration in solution. For IBS1* this is $k_{on,01} = k'_{on} \times c^{-1}_{IBS1*}$.

[b] Dissociation constants, K_d, are calculated from the logistic fits (Eqn. S13) to the cumulative distribution of $K_{d,n}$ values (single-molecule) of directly from the rates $K_d = k_{off} \times k_{on}^{-1}$ (SPR).

[c] Value from one SPR sensor chip (no replicate).

Table 4.4 | Kinetic parameters from cumulative dwell time histogram analysis of single-molecule FRET experiments.

Cognate strand	Metal ion concentration	k_{on} ($10^5\,M^{-1}\,s^{-1}$) [a]	β	a_1	$k_{off,1}$ (s^{-1})	$k_{off,2}$ (s^{-1})	$K_{d,1}$ (nM)	$K_{d,2}$ (nM)	$\Delta G_{bind,1}$ ($kJ\,mol^{-1}$)	$\Delta G_{bind,2}$ ($kJ\,mol^{-1}$)	$\Delta^{\ddagger}$ ($kJ\,mol^{-1}$)
IBS1*	1 M K^+ [b]	8.84 ± 0.22	0.99 ± 0.01	0.10 ± 0.02	0.054 ± 0.002	–	1090 ± 320	–	-34.0 ± 0.7	–	39.0 ± 0.1
IBS1*	0 mM Mg^{2+}	3.97 ± 0.64	1.00 ± 0.01	0.38 ± 0.08	1.1 ± 0.4	0.055 ± 0.011	2960 ± 113	143(39)	-32 ± 1	-39.0 ± 0.7	41.0 ± 0.4
IBS1*	0.5 mM Mg^{2+}	4.85 ± 0.51	0.89 ± 0.07	0.43 ± 0.06	0.49 ± 0.12	0.021 ± 0.003	1020 ± 250	44(7)	-34.0 ± 0.6	-42.0 ± 0.4	41.0 ± 0.3
IBS1*	2.5 mM Mg^{2+}	4.29 ± 0.35	0.95 ± 0.05	0.57 ± 0.05	0.31 ± 0.04	0.012 ± 0.002	738 ± 112	29(4)	-35.0 ± 0.4	-43.0 ± 0.4	41.0 ± 0.2
IBS1*	5 mM Mg^{2+}	5.71 ± 0.61	0.84 ± 0.04	0.55 ± 0.03	0.35 ± 0.05	0.012 ± 0.002	619 ± 104	20(4)	-35.0 ± 0.4	-44.0 ± 0.4	40.0 ± 0.3
IBS1*	10 mM Mg^{2+}	9.64 ± 0.82	0.88 ± 0.04	0.64 ± 0.04	0.28 ± 0.03	0.013 ± 0.002	5291 ± 40	13(2)	-37.0 ± 0.3	-45.0 ± 0.4	39.0 ± 0.2
IBS1*	15 mM Mg^{2+}	7.20 ± 0.92	0.89 ± 0.07	0.48 ± 0.08	0.36 ± 0.12	0.012 ± 0.002	502 ± 163	17(3)	-36.0 ± 0.8	-44.0 ± 0.5	40.0 ± 0.3
IBS1*	20 mM Mg^{2+}	8.06 ± 0.69	0.93 ± 0.07	0.39 ± 0.06	0.43 ± 0.13	0.012 ± 0.002	541 ± 159	16(3)	-36.0 ± 0.7	-45.0 ± 0.4	39.0 ± 0.2
dIBS1*	1 M K^+ [d]	3.79 ± 1.25	0.77 ± 0.10	1.00 ± 0.00	3.48 ± 0.32	–	$10^4 \pm 3200$	–	-29.0 ± 0.8	–	41.0 ± 0.8
dIBS1*	2.5 mM Mg^{2+}	8.03 ± 0.66	0.71 ± 0.02	0.95 ± 0.02	0.80 ± 0.04	0.098 ± 0.030	1004 ± 104	122 ± 39	-34.0 ± 0.3	-39.0 ± 0.8	39.0 ± 0.2
dIBS1*	5 mM Mg^{2+}	8.76 ± 0.29	0.79 ± 0.01	0.94 ± 0.01	0.65 ± 0.01	0.071 ± 0.007	747 ± 27	82 ± 8	-35.0 ± 0.1	-40.0 ± 0.3	39.0 ± 0.1
dIBS1*	10 mM Mg^{2+}	10.6 ± 0.3	0.85 ± 0.01	0.94 ± 0.01	0.57 ± 0.01	0.064 ± 0.006	540 ± 16	60 ± 6	-36.0 ± 0.1	-41.0 ± 0.3	39.0 ± 0.1
dIBS1*	15 mM Mg^{2+}	12.4 ± 0.3	0.89 ± 0.01	0.95 ± 0.01	0.48 ± 0.01	0.059 ± 0.005	389 ± 13	48 ± 5	-37.0 ± 0.1	-42.0 ± 0.2	38.0 ± 0.1
dIBS1*	20 mM Mg^{2+}	14.8 ± 0.1	0.92 ± 0.01	0.96 ± 0.01	0.48 ± 0.01	0.051 ± 0.004	325 ± 6	35 ± 3	-37.0 ± 0.1	-43.0 ± 0.2	38.0 ± 0.1

Errors are standard deviations of 100 bootstrap samples.[320]
All experiments were conducted with a monovalent ionic background of 100 mM KCl unless stated otherwise.
[a] Second-order rate coefficients, $k_{on,1}$, are calculated from the pseudo-first-order association rate coefficient and the total IBS1* concentration in solution (35 nM IBS1* or 50 nM dIBS1*). For IBS1* this is $k_{on,1} = k'_{on,1} \times c^{-1}_{IBS1*}$.
[b] 100 µM EDTA are added to chelate any traces of divalent metal ions.

Table 4.5 | Kinetic parameters determined with a trained global hidden Markov model from single-molecule FRET experiments.

Cognate strand	Metal ion concentration	$k'_{on,01}$ ($10^{-2}\,s^{-1}$) [a]	$k_{on,01}$ ($10^{5}\,M^{-1}\,s^{-1}$) [b]	$k_{off,10}$ (s^{-1})	k_{12} ($10^{-2}\,s^{-1}$)	k_{21} ($10^{-2}\,s^{-1}$)	K_d (nM)	ΔG_{bind} ($kJ\,mol^{-1}$)	$\Delta^{\ddagger}$ ($kJ\,mol^{-1}$)
IBS1*	1 M K^+ [b]	2.96 ± 0.04	8.45 ± 0.11	0.051 ± 0.001	–	–	59.0	-41.2	39.2 ± 0.1
IBS1*	0 mM Mg^{2+}	1.69 ± 0.07	4.82 ± 0.14	0.136 ± 0.010	–	–	281	-37.4	40.6 ± 0.1
IBS1*	0.5 mM Mg^{2+}	1.08 ± 0.04	3.08 ± 0.10	0.351 ± 0.016	47.1 ± 2.4	4.28 ± 0.19	81.5	-40.4	41.7 ± 0.1
IBS1*	2.5 mM Mg^{2+}	1.02 ± 0.03	1.92 ± 0.08	0.177 ± 0.007	15.0 ± 0.7	1.52 ± 0.06	46.0	-41.9	41.8 ± 0.1
IBS1*	5 mM Mg^{2+}	1.44 ± 0.03	4.10 ± 0.09	0.179 ± 0.006	21.4 ± 0.7	1.70 ± 0.05	28.5	-43.0	41.0 ± 0.1
IBS1*	10 mM Mg^{2+}	2.25 ± 0.05	6.44 ± 0.15	0.151 ± 0.005	15.2 ± 0.6	1.23 ± 0.04	17.2	-44.3	39.8 ± 0.1
IBS1*	15 mM Mg^{2+}	2.05 ± 0.07	5.87 ± 0.19	0.118 ± 0.005	20.4 ± 1.0	2.21 ± 0.09	17.5	-44.3	40.1 ± 0.1
IBS1*	20 mM Mg^{2+}	2.14 ± 0.07	6.11 ± 0.20	0.146 ± 0.006	33.1 ± 1.4	2.56 ± 0.10	15.2	-44.6	40.0 ± 0.1
dIBS1*	1 M K^+ [d]	0.86 ± 0.03	1.72 ± 0.06	3.52 ± 0.11	–	–	17490	-27.2	43.1 ± 0.1
dIBS1*	2.5 mM Mg^{2+}	1.87 ± 0.03	3.75 ± 0.06	0.740(13)	3.83 ± 0.35	6.68 ± 0.53	1190	-33.8	41.2 ± 0.1
dIBS1*	5 mM Mg^{2+}	3.47 ± 0.03	6.94 ± 0.05	0.619(5)	2.78 ± 0.14	5.76 ± 0.25	582	-35.6	39.7 ± 0.1
dIBS1*	10 mM Mg^{2+}	5.15 ± 0.03	10.3 ± 0.1	0.585(4)	2.40 ± 0.10	2.66 ± 0.11	306	-37.2	38.7 ± 0.1
dIBS1*	15 mM Mg^{2+}	7.01 ± 0.05	14.0 ± 0.1	0.508(4)	3.46 ± 0.13	7.85 ± 0.24	244	-37.7	37.9 ± 0.1
dIBS1*	20 mM Mg^{2+}	7.07 ± 0.03	14.1 ± 0.1	0.511(2)	2.35 ± 0.07	6.33 ± 0.16	255	-37.6	37.9 ± 0.1

Errors are standard deviations computed from likelihood ratio tests as detailed in ref. [318].
All experiments were conducted with a monovalent ionic background of 100 mM KCl unless stated otherwise. [a] Pseudo-first-order association rate coefficients, $k'_{on,01}$, are extracted from the HMM.

[b] Second-order rate coefficients, $k_{on,01}$, are calculated from the pseudo-first-order association rate coefficient and the total IBS1* concentration in solution (35 nM IBS1* or 50 nM dIBS1*). For IBS1* this is $k_{on,01} = k'_{on} \times c^{-1}_{IBS1*}$.
[c] Dissociation constants are calculated from the relative state population S_{bound} and $S_{unbound}$ and the ligand concentration. For IBS1* the K_d is given by $S_{unbound} \times c_{IBS1*} \times S^{-1}_{bound}$. [d] 100 µM EDTA are added to chelate any traces of divalent metal ions.

Table 4.6 | Comparison of kinetic parameters determined from dwell time fits to the experimental decays (denoted as "exp.", Supplementary Table 4.4) with the dwell time fits after resimulating time traces from the rate system determined by the global hidden Markov model (denoted as "sim.").

Cognate strand	**Metal ion concentration**	k'_{on} ($10^{-2}\,s^{-1}$, exp.) [a]	k'_{on} ($10^{-2}\,s^{-1}$, sim.) [a]	$k_{off,1}$ (s^{-1}, exp.)	$k_{off,1}$ (s^{-1}, sim.)	$k_{off,2}$ (s^{-1}, exp.)	$k_{sim,2}$ (s^{-1}, exp.)	a_1 (exp.)	a_2 (sim.)
IBS1*	1 M K^+ [b]	3.1(1)	3.1(2)	0.054(2)	0.056(3)	–	–	1	1
IBS1*	20 mM Mg^{2+}	2.8(2)	3.0(4)	0.43(13)	0.19(9)	0.013(2)	0.014(2)	0.39(6)	0.22(8)
dIBS1*	1 M K^+ [b]	1.9(6)	1.4(2)	3.48(32)	3.33(33)	–	–	1	1
dIBS1*	20 mM Mg^{2+}	7.4(1)	7.4(2)	0.48(1)	0.52(2)	0.052(4)	0.018(4)	0.96(1)	0.96(1)

Errors are standard deviations of 100 bootstrap samples.[320]
Experiments at 20 mM Mg^{2+} were conducted with a monovalent ionic background of 100 mM KCl.
[a] Pseudo-first-order association rate coefficients, $k'_{on,01}$.
[b] 100 µM EDTA are added to chelate any traces of divalent metal ions.

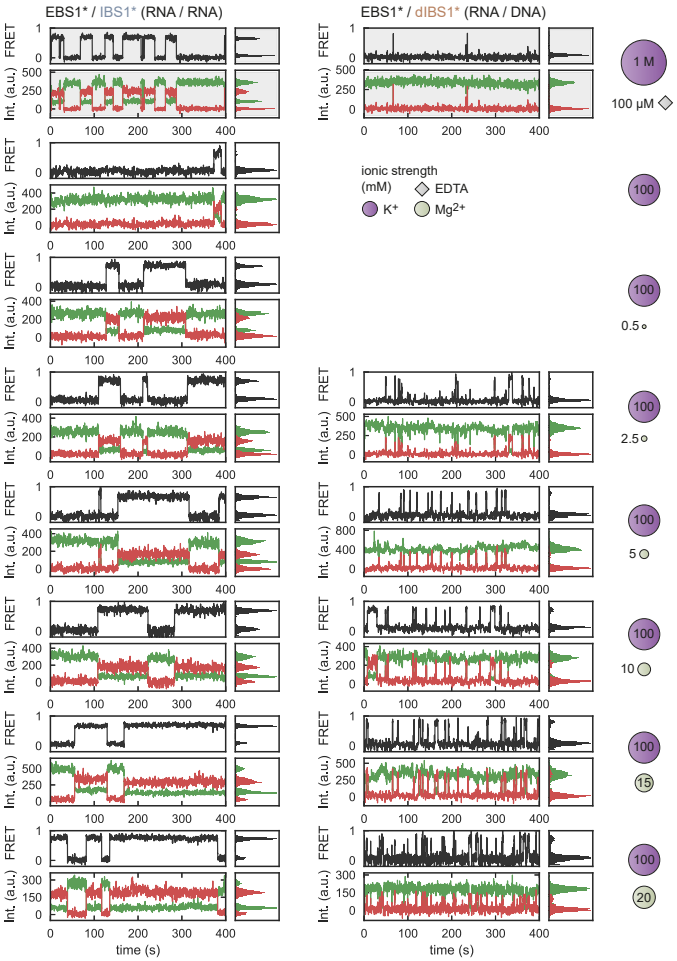


Fig. 4.6 | Dynamic single-molecule intensity and FRET time traces of EBS*/IBS1* (left) and EBS1/dIBS1* (right) at different Mg^{2+} concentrations. Anticorrelated Cy3 (green) and Cy5 (red) signals translate into FRET efficiencies that interconvert between a zero (unbound) and high (~ 0.75, bound) state. The traces are selected to reflect the thermodynamic state population in the overall histograms in Fig. 4.1f. Source data are provided as a Source Data file.

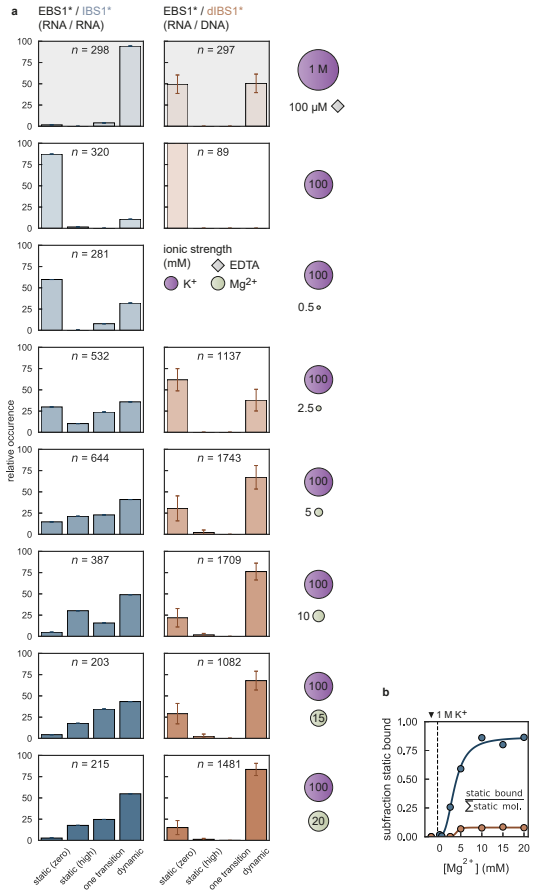


Fig. 4.7 | Classification of molecules according to their interconversion frequency between zero and high FRET states. **a** Bar charts depict relative abundance of static and dynamic molecules for EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange). Error bars are drawn from three independent experiments (for EBS1*/dIBS1* only) and correspond to mean ± s.d. The total number of molecules n is indicated for each condition. Mg^{2+} concentration is varied across the datasets (0 mM–20 mM) under constant background of 100 mM KCl. High monovalent salt conditions (1 M K^+, 100 µM EDTA) are represented in the top row (gray background). **b** Subfraction of statically bound molecules relative to all static molecules across different Mg^{2+} concentrations. Source data are provided as a Source Data file.

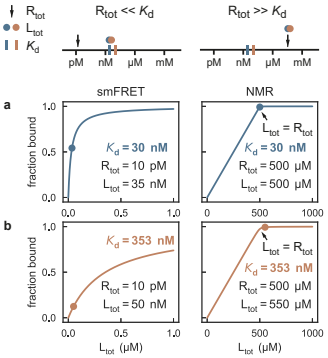


Fig. 4.8 | Binding isotherms for a 1:1 interaction between a receptor R and a ligand L. The curves represent the fraction of bound molecules as a function of the total ligand concentration as obtained from equation 4.32. The isotherm is computed for two sets of receptor concentrations: $[R_{\mathrm{tot}}] = 10\,\mathrm{pM}$ (the concentration of immobilized EBS1* in the smFRET experiments) and $[R_{\mathrm{tot}}] = 500\,\mu\mathrm{M}$ (the minimum concentration of EBS1* at which the NMR structures are solved). The dissociation constants characterizing the binding affinity of the ligand to the receptor are either (a) 30 nM (EBS1*/IBS1*, blue) or (b) 353 nM (EBS1*/dIBS1*, orange). The circles indicate the ligand concentration at which the experiment is recorded. In the single-molecule TIRF experiments the hairpin is present at concentrations far below the dissociation constant: $[R_{\mathrm{tot}}] \ll K_{\mathrm{d}}$. Due to background fluorescence the ligand concentration is limited to about 50 ¡nanoM, which is not high enough to saturate the receptor. In the NMR tube $[R_{\mathrm{tot}}] \gg K_{\mathrm{d}}$ and the isotherm shows a sharp kink where $[L_{\mathrm{tot}}] = [R_{\mathrm{tot}}]$ (black arrow). Here, at equimolar concentrations of ligand and receptor, EBS1* is fully saturated.

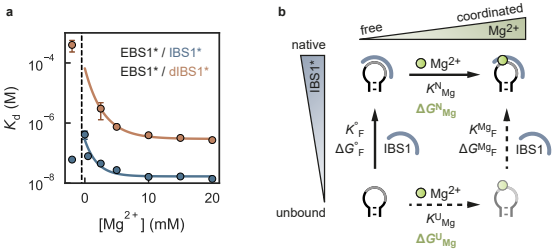


Fig. 4.9 | Thermodynamics of tertiary contact formation. **a** Dissociation constants of EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange) as a function of Mg^{2+} calculated from the relative population of the zero and high FRET state in the FRET histograms from dynamic molecules. The affinity of the dIBS1* towards the hairpin is more than one order of magnitude lower across all Mg^{2+} concentrations. Error bars are drawn as mean ± s.d. of 100 bootstrap samples. **b** Thermodynamic circle linking tertiary contact formation and Mg^{2+} binding. The transition from the unbound (U) to the native state (N) is described by the equilibrium constants K°_F in the absence and $K^{Mg^{2+}}_F$ in the presence of Mg^{2+}. The free energy contribution of Mg^{2+} to tertiary contact formation is given $\Delta\Delta G_{Mg^{2+}} = \Delta G^N_{Mg^{2+}} - \Delta G^U_{Mg^{2+}} = \Delta G^F_{Mg^{2+}} - \Delta G^\circ_F$.[296] Binding of Mg^{2+} to the unbound state (U • Mg^{2+}, grayed out) is unlikely to be observed since a discrete binding pocket is only formed when IBS1* has bound to the EBS1* hairpin (Fig. 4.3b). This state and its connections are thus omitted in the HMM model and the thermodynamic cycle is transformed into a linear three-state model.

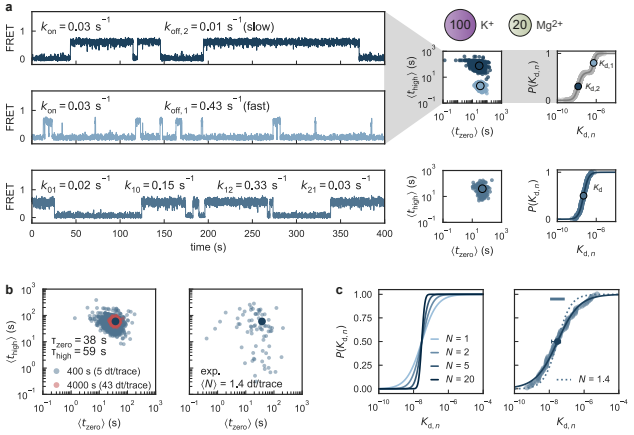


Fig. 4.10 | Simulation of slow exchange between kinetic regimes leading to intermolecule heterogeneity. **a** Simulated homogenous two-state systems with a slow ($k_{\mathrm{off},2}$, top) or fast ($k_{\mathrm{off},2}$, middle) dissociation rate as obtained from dwell time analysis (Supplementary Table 4.4 and Supplementary Fig. 4.14) of single-molecule FRET experiments at 100 mM KCl and 20 mM $MgCl_2$. Upon merging both simulations, the mean dwell-time plot split into two well separated clusters due to the separation of timescales of the simulated rates and because, by design, the two species do not interconvert. Thus, the cumulative distribution of all $K_{\mathrm{d},n}$ shows two transitions one for each dissociation constant, $K_{\mathrm{d},1}$ and $K_{\mathrm{d},2}$. Conversely, if the rate system allows interconversion (k_{12} and k_{21}, bottom) between the kinetic regimes (see HMM model, Supplementary Table 4.5 and Supplementary Fig. 4.16) the mean dwell time is averaged and only a single, stretched cluster remains. Here, short and long binding events occur in the same trace and the dissociation constant, K_{d}, is calculated from a single transition. **b** Effect of the average number of dwell times $\langle N \rangle$ per trace on the spread of the mean dwell time distribution. The number of dwell times N for an individual trace is computed as $(i_{\max} + j_{\max})/2$. A two-state system with the experimentally determined dwell time parameters $\tau_{\mathrm{zero}} = 38\,\mathrm{s}$ and $\tau_{\mathrm{high}} = 59\,\mathrm{s}$ is simulated with trace lengths of 400 s (blue, experimental observation time) and 4000 s (red) resulting in 5 or 43 dwell times per trace on average. For comparison, the experimental mean dwell time scatter plot is shown for EBS1*/IBS1 at 20 mM Mg^{2+}. **c** The average number of dwell times per trace affects the steepness of the $K_{\mathrm{d},n}$ distribution. $P(K_{\mathrm{d},n})$ is calculated as the numerical integral over equation 4.17 (i.e. the cumulative sum over the ratio of two gamma distributions) for different $N = i_{\max} = j_{\max}$. The experimental distribution is compared to the numerical integral for $N = 1.4$.

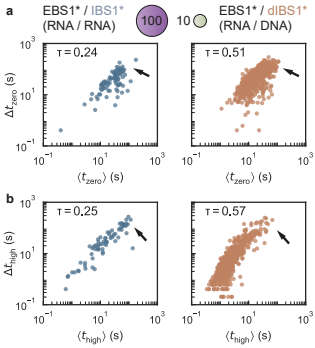


Fig. 4.11 | Correlation of mean dwell time $\langle t_{\mathrm{high}} \rangle$ or $\langle t_{\mathrm{zero}} \rangle$ with the difference between the shortest and longest dwell time Δt_{high} or Δt_{zero} for EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange). Kendall's tau coefficient is indicated for each plot as a measure for the correlation. The arrow points to the region of the most heterogeneous molecules (large Δt).

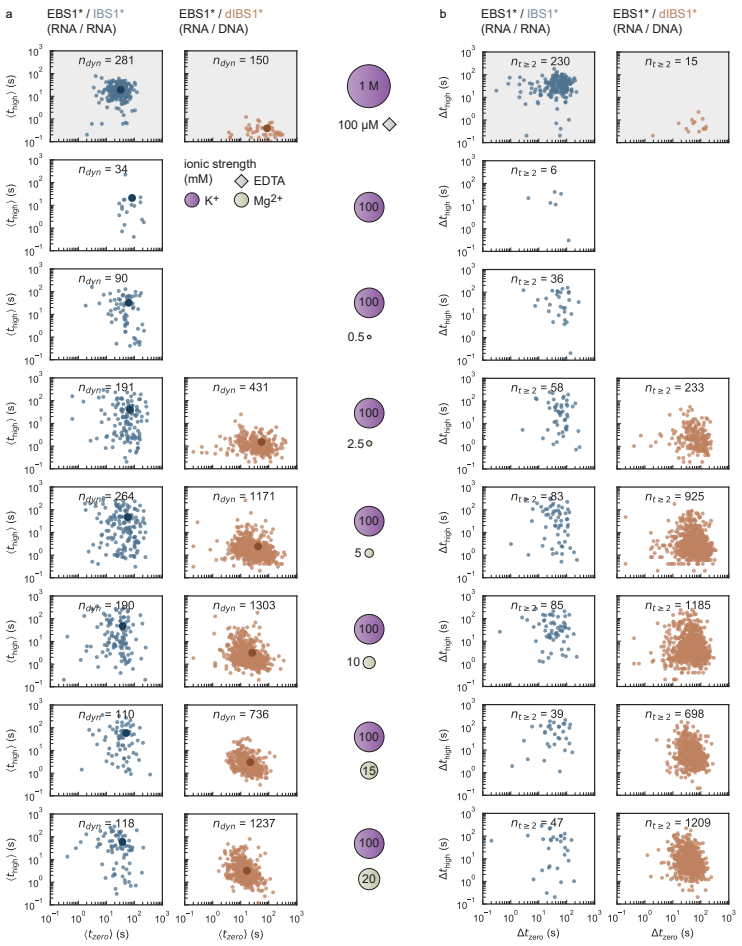


Fig. 4.12 | Experimentally characterized molecule-to-molecule variations in EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange). **a** Mean dwell time correlation plots of the bound, $\langle t_{\mathrm{high}} \rangle$, and unbound state, $\langle t_{\mathrm{zero}} \rangle$ across different Mg^{2+} concentrations. High monovalent salt conditions (1 M K^+, 100 µM EDTA) are represented in the top row (gray background). The cluster center drawn as a circle and the number of dynamic molecules n_{dyn} is indicated for each plot. **b** Difference between the shortest and longest dwell time within a trace for the bound Δt_{high} and unbound Δt_{zero} state. The number of dynamic molecules with at least two dwell times $n_{t \geq 2}$ are indicated for each plot. Source data are provided as a Source Data file.

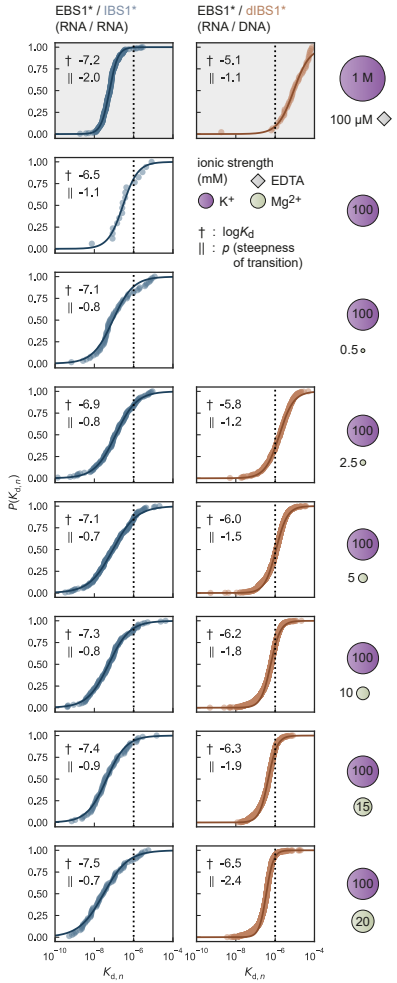


Fig. 4.13 | Normalized cumulative $K_{\mathrm{d},n}$ histograms for EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange) across different Mg^{2+} concentrations. The top row refers to a high monovalent ionic atmosphere in the absence of divalent metal ions (1 M K^+, 100 µM EDTA). A dotted vertical line is drawn at $K_\mathrm{d} = 10^{-6}$ M as a guide for the eye. A homogeneous logistic model (Eqn. 4.19) is fitted to the cumulative histograms. The sigmoid is characterized by two parameters, the midpoint K_d and the steepness of the curve p, which are indicated alongside each plot.

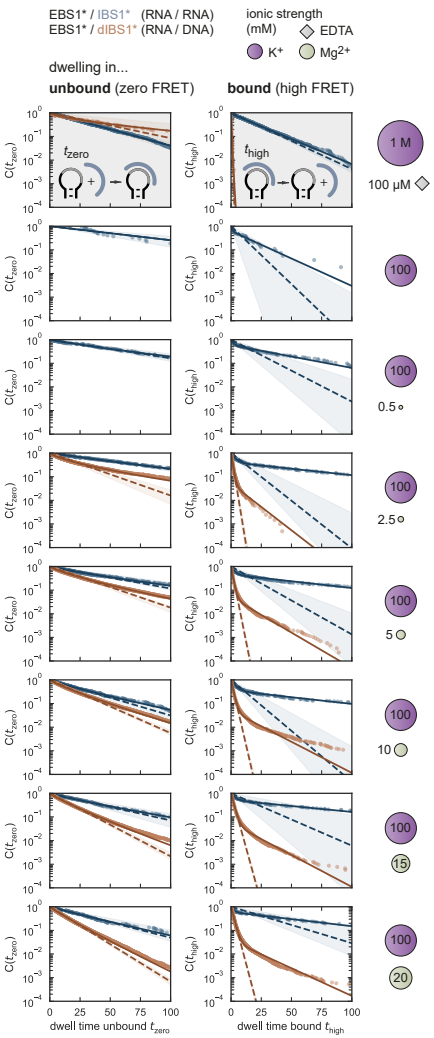


Fig. 4.14 | Complementary cumulative dwell time distributions of the unbound (left) and bound (right) state for EBS1*/IBS1* (blue) and EBS1/dIBS1* (orange) across different Mg^{2+} concentrations. The complementary cumulative distribution $C(t_{\mathrm{zero}})$ and and $C(t_{\mathrm{high}})$ represent the probability of a molecule to survive beyond a time t_{zero} or thigh in the given state. The characteristic dwell time decay constants τ_{zero} and τ_{high} are obtained by fitting equation 4.8 (for τ_{high}) or equation 4.9 (for τ_{zero}) to the distributions. (*continued on next page*)

Fig. 4.14 | *(continued)* Single-exponential fits (homogeneous binding/unbinding kinetics) are drawn as dashed lines. They don't account for the heterogeneous unbinding kinetics in either of the two systems. The 0.99 bootstrap confidence interval is indicated by shading. The logarithmic scaling emphasizes longer dwell times also if they are low-abundant. Importantly, some heterogeneity persists even in the RNA/DNA hybrids, yet to a much lower extent that in the RNA/RNA contact. The top row refers to a high monovalent ionic atmosphere in the absence of divalent metal ions (1 M K^+, 100 µM EDTA). Source data are provided as a Source Data file.

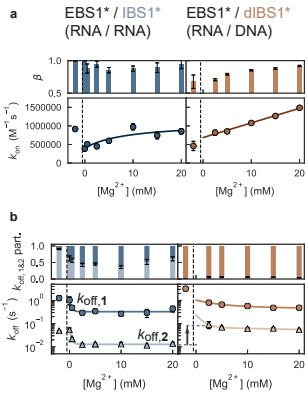


Fig. 4.15 | Binding and unbinding rates from complementary cumulative dwell time histograms for EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange) across different Mg^{2+} concentrations. **a** Second-order binding rates, kon, are obtained from fitting a stretched single-exponential function (Eqn. 4.9) to the complementary cumulative unbound dwell time histograms in Supplementary Fig. 4.14 and taking into account the total IBS1* or dIBS1* concentration. The stretching factor β is indicated in a bar chart and represents slight inhomogeneities in the binding rates. **b** Unbinding rates $k_{\mathrm{off},1}$ (dark blue/orange) and $k_{\mathrm{off},2}$ (light blue/orange), are plotted as circles or triangles respectively. The partition coefficients (weighting factors) of the two off-rates are drawn as stacked bars. Errors are given as mean $\pm$ s.d. of 100 bootstrap samples.

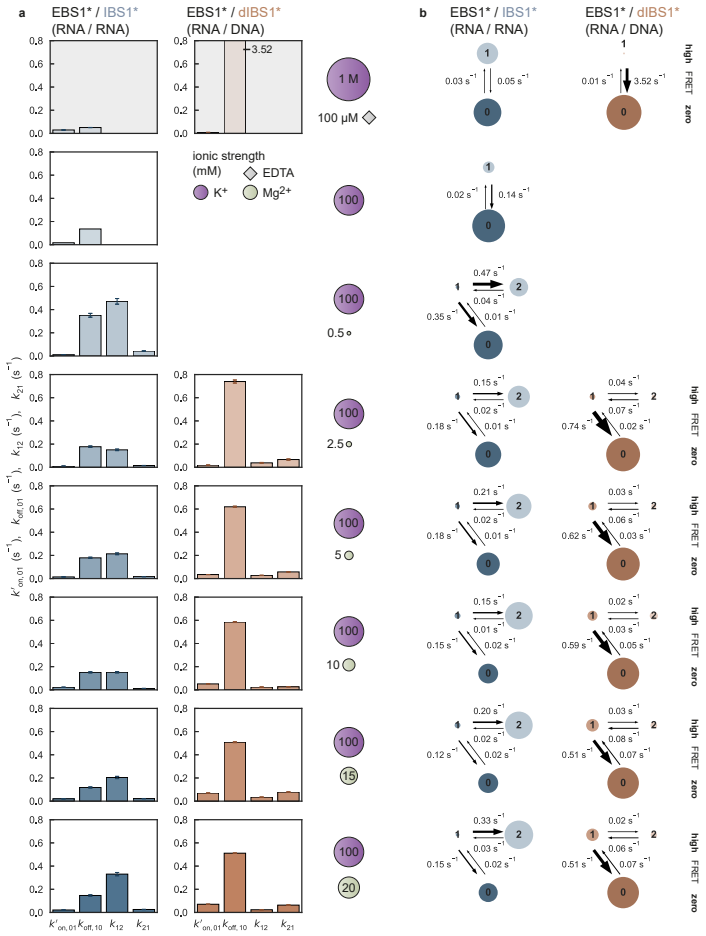


Fig. 4.16 | Binding and unbinding rates from a trained global hidden Markov model (HMM) for EBS1*/IBS1* (blue) and EBS1*/dIBS1* (orange) across different Mg^{2+} concentrations. **a** Bar chart of the interconversion rates $k'_{\mathrm{on},01}$, $k_{\mathrm{off},10}$, k_{12} and k_{21} obtained from training an HMM on all dynamic FRET traces. **b** Two-state (top row, 1 M K^+, 100 µM EDTA) and three-state models at different Mg^{2+} concentrations. The relative state population is proportional to the area of the circle and the magnitude of the (pseudo)first order rate coefficients is linked to the arrow width. In the three-state models the high FRET state is two-fold degenerate, that is state 1 (exon bound, Mg^{2+} unbound) and 2(exon bound, Mg^{2+} bound) have the same FRET efficiency and can only be differentiate kinetically. Errors are standard deviations calculated from likelihood ratio tests as detailed in ref. [318]. Source data is provided in Supplementary Table 4.5.

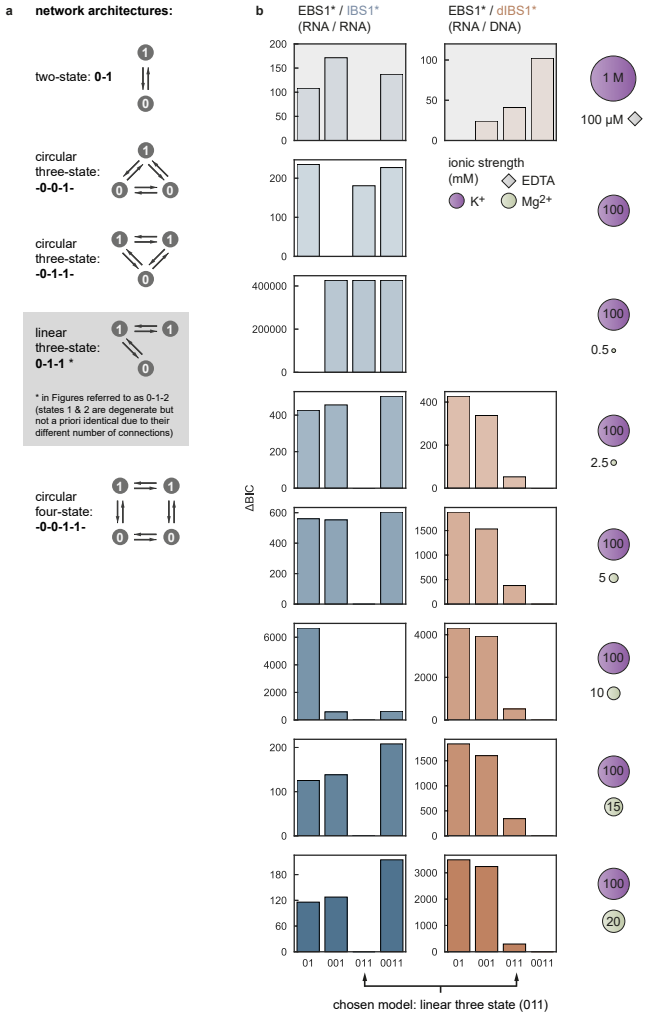


Fig. 4.17 | Bayesian Information criterion (BIC) of trained hidden Markov models. **a** Different network architectures of the tested HMMs. **b** Bar plots show ΔBIC values with respect to the best model where the absolute BIC value is minimized. Since the number of states is determinant for the criterion, the circular and linear three state model 011 have the same BIC value. The BIC is used in combination with biochemical knowledge about the system to select the most appropriate model (linear three-state: 0-1-1) which is shaded in gray.

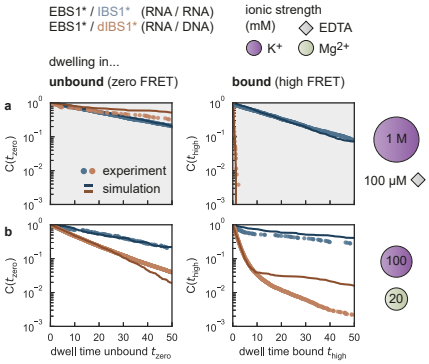


Fig. 4.18 | Dwell time distributions reproduced from simulation of the HMM derived rate system (linear three-state model, Supplementary Fig. 4.16). Simulated (lines) and experimental (dots) decays are overlaid for **a** 1 M K^+ and 100 µM EDTA and **b** 100 mM K^+ and 20 mM Mg^{2+}. Experimental and simulated distributions are in good agreement, thus supporting the chosen kinetic model.

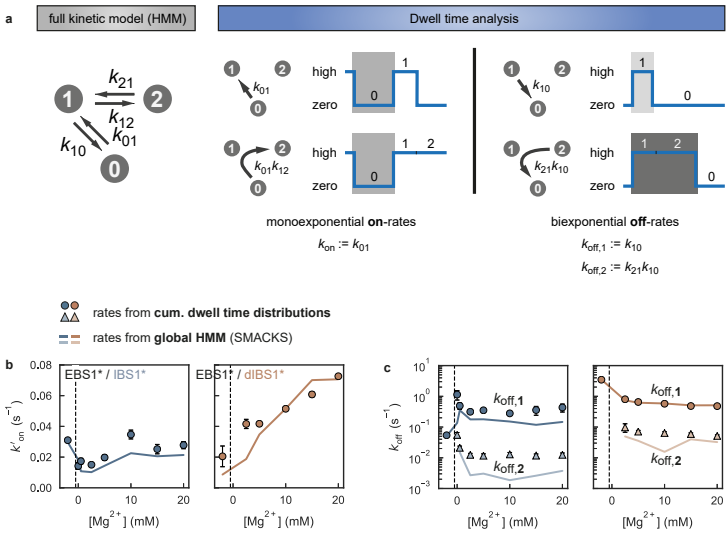


Fig. 4.19 | Comparison of dwell time analysis with hidden Markov modeling (HMM). **a** The HMM includes interconversion rates k_{12} and k_{21} between the degenerated kinetic states 1 and 2. Dwell time analysis of the FRET states (bound/unbound) is blind to this interconversion. Hence, the observed on-rate is mono-exponential. Dissociation from the bound state can occur from either kinetic state 1 or 2 and therefore lead to a bi-exponential decay where $k_{\mathrm{off},1}$ is equal to k_{10} and $k_{\mathrm{off},2}$ is the product of k_{21} and k_{10}. **b/c** Comparison of kinetic rates determined from fits to cumulative dwell time distributions (markers) and extracted from the trained global hidden Markov model (lines) as a function of the Mg^{2+} concentration.

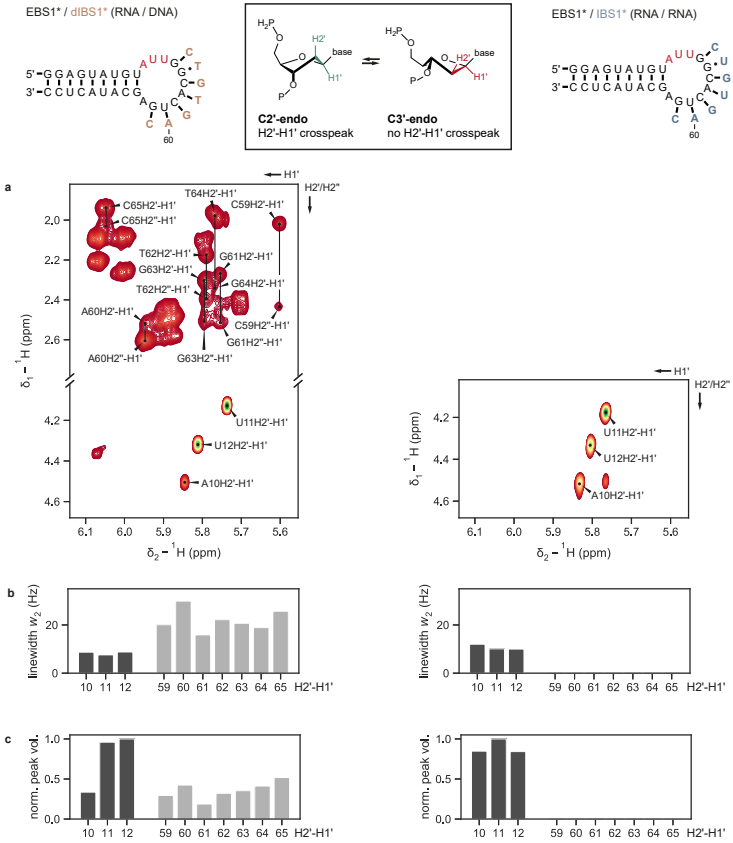


Fig. 4.20 | Weak H2′-H1′ coupling and broad linewidths suggest exchange between sugar pucker conformations in dIBS1*. **a** [^{1}H,^{1}H]-TOCSY spectrum of EBS1*/dIBS11* and EBS1*/IBS1*. In a sugar C2′-endo conformation, coupling between H2′ and H1′ gives rise to a strong cross-peak as seen for nucleotides U11 and U12 (marked in red for both EBS1*/dIBS11* and EBS1*/IBS11*). In a C3′-endo sugar the H2′ and H1′ protons are nearly perpendicular to each other, which weakens the ^{3}J-coupling constant and no cross peak is observed (EBS1*/IBS1*). **b** Linewidths and **c** peaks volumes are calculated by 2D-Gaussian fits. Broad line shapes and intermediary peak volumes for H2′-H1′ correlations of all residues in dIBS1*, suggesting fast interconversion between sugar puckers (spectrum of EBS1*/IBS1* adapted from ref. [292]). On the contrary, IBS1* shows no H2′-H1′ correlations (H2′ chemical shifts in RNA usually lie in the range of 4 ppm)–5 ppm), which points to a stable C3′-endo conformation. Spectra were recorded at 298 K in D_2O in the presence of 110 mM KCl.

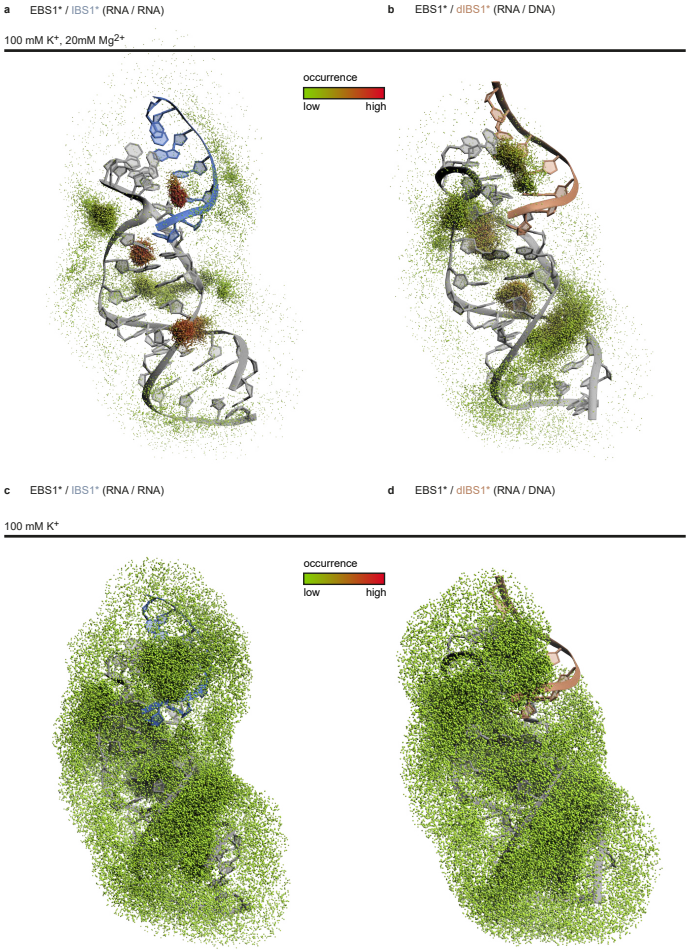


Fig. 4.21 | Spatial distribution of K^+ and Mg^{2+} binding from MD simulations. All ion positions (Mg^{2+} for a/b; K^+ for c/d) within 6 Å of the nucleic acids are color-coded based on their occurrence (highest: red, lowest: green). The colors are scaled to the same probability density across all four simulations. **a/b** Mg^{2+} binding based on experimentally observed coordination sites (NOEs to $[Co(NH_3)_6]^{3+}$, Mg^{2+} induced chemical shifts) used as starting positions for the MD simulations. **c/d** K^+ binds much more uniformly to EBS1*/IBS1* and EBS1*/dIBS1*. The point cloud encompasses loosely associated, outer-sphere and inner-sphere bound Mg^{2+} ions.

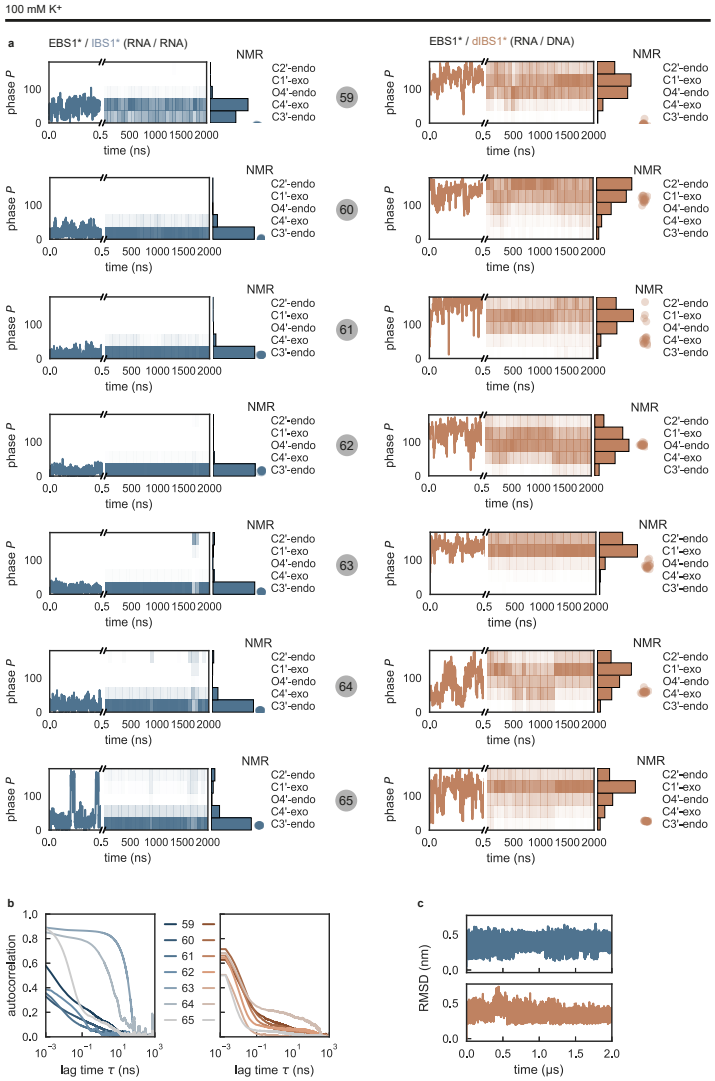


Fig. 4.22 | Sugar pucker dynamics of IBS1* and dIBS1* from MD simulations with 100 mM K^+ only. **a** Time evolution of the pseudorotation (phase) angle for all residues in IBS1* (blue) or dIBS1* (orange). **b** Autocorrelation of the pseudorotation angle for each residue of IBS1* or dIBS1*. **c** RMSD of the RNA dynamics along the trajectory.

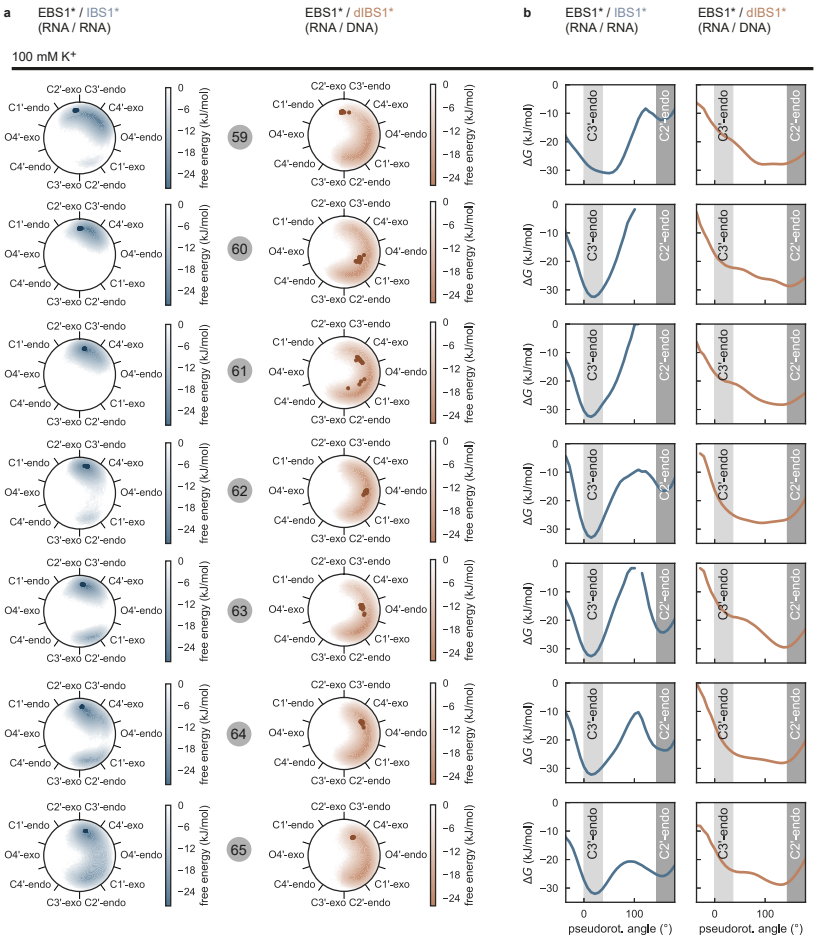


Fig. 4.23 | Sugar pucker conformations of IBS1* and dIBS1* from MD simulations in the absence of Mg^{2+} but with 100 mM K^+ only. **a** Pseudorotation cycles for all residues of IBS1* or dIBS1* (C59-C65). Distributions of pucker phase and amplitude are color coded by their free energy. The puckers of the 18 (EBS1*/IBS1*, blue) or 20 (EBS1*/dIBS1*, orange) lowest energy NMR structures are indicated by dots. **b** Free energy profiles of the pseudorotation angle (pucker phase) for all residues of IBS1* or dIBS1*.

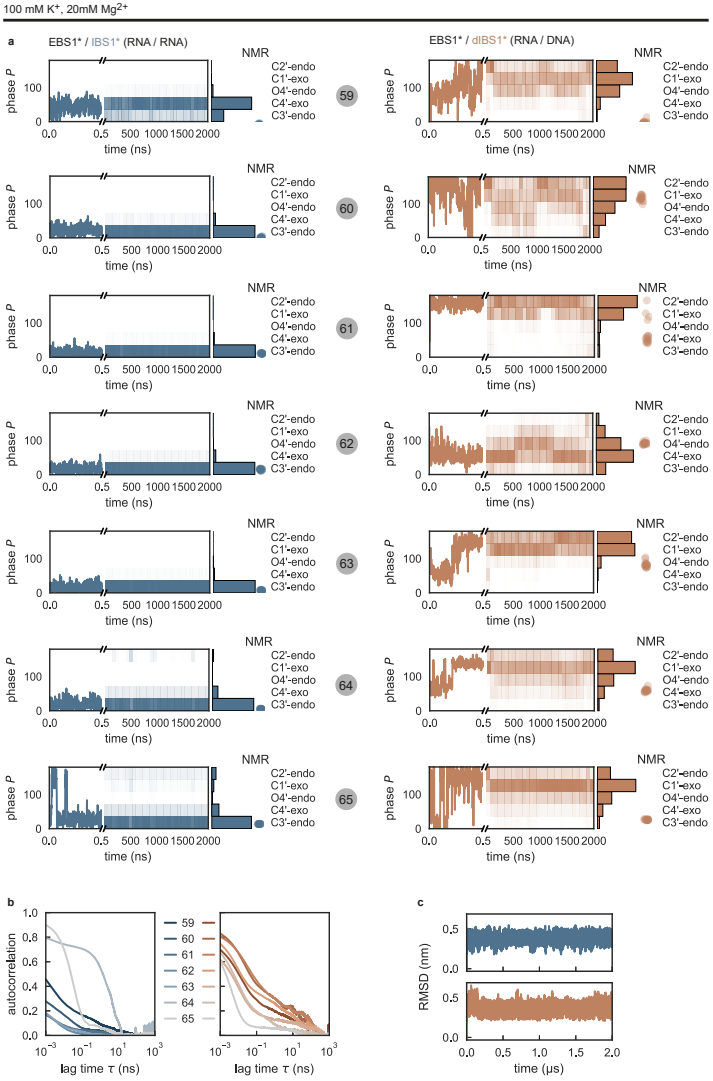


Fig. 4.24 | Sugar pucker dynamics of IBS1* and dIBS1* from MD simulations in the presence of 100 mM K^+ and 20 mM Mg^{2+}. **a** Time evolution of the pseudorotation (phase) angle for all residues in IBS1* (blue) or dIBS1* (orange). **b** Autocorrelation of the pseudorotation angle for each residue of IBS1* or dIBS1*. **c** RMSD of the RNA dynamics along the trajectory.

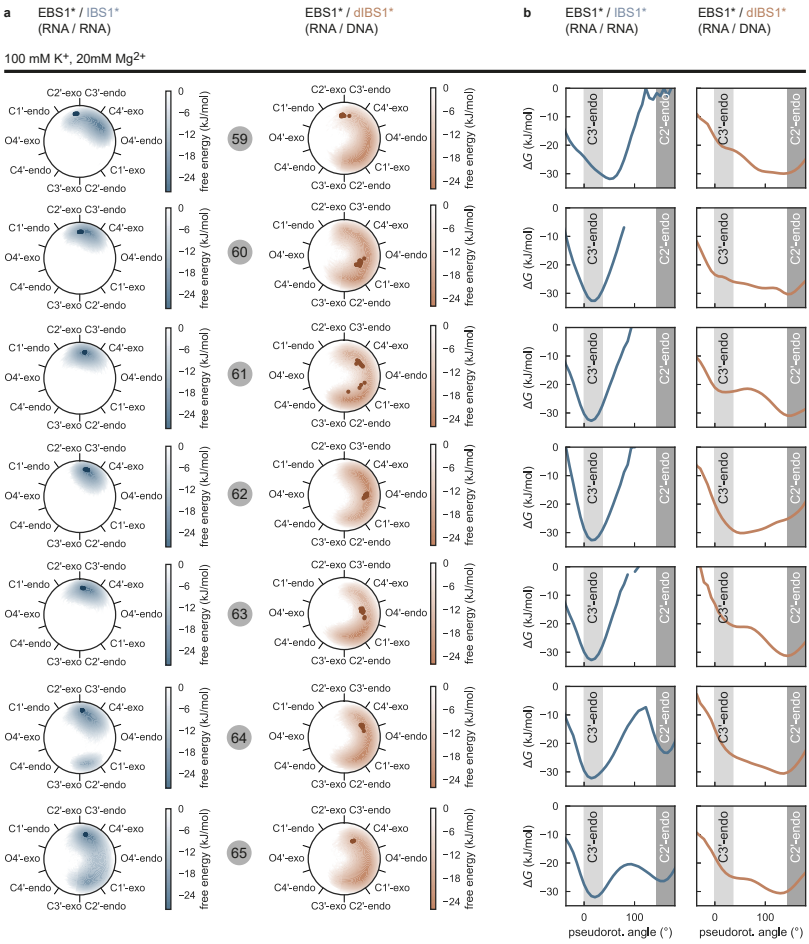


Fig. 4.25 | Sugar pucker conformations of IBS1* and dIBS1* from MD simulations with 100 mM K^+ and 20 mM Mg^{2+}. **a** Pseudorotation cycles for all residues of IBS1* or dIBS1* (C59-C65). Distributions of pucker phase and amplitude are color coded by their free energy. The puckers of the 18 (EBS1*/IBS1*, blue) or 20 (EBS1*/dIBS1*, orange) lowest energy NMR structures are indicated by dots. **b** Free energy profiles of the pseudorotation angle (pucker phase) for all residues of IBS1* or dIBS1*.

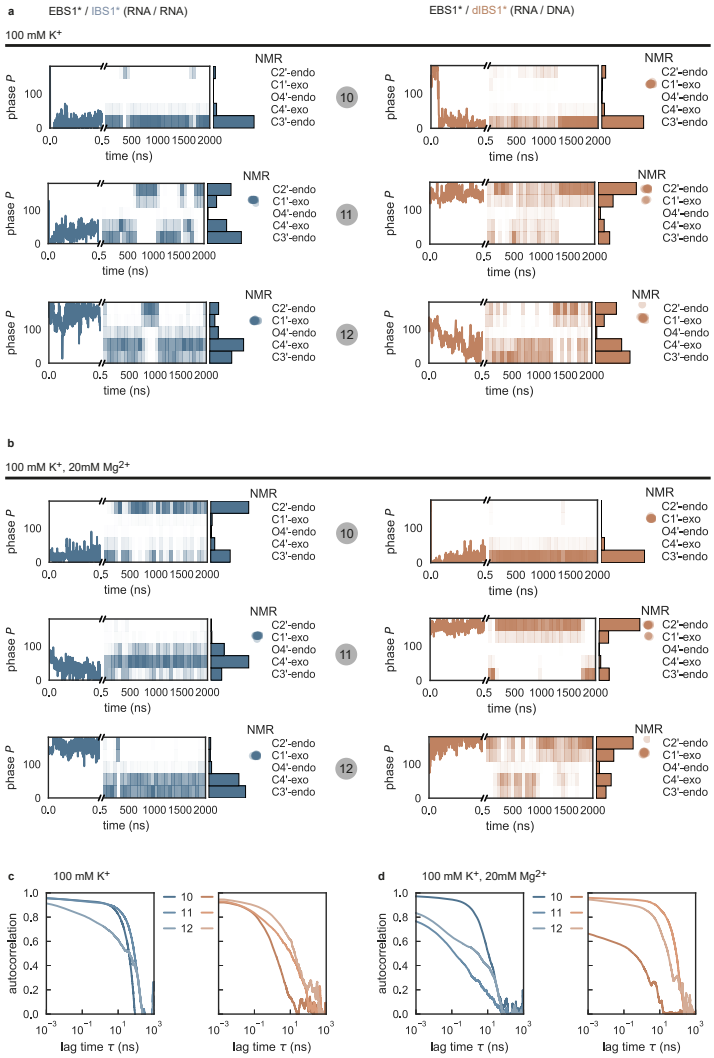


Fig. 4.26 | Sugar pucker dynamics of EBS1* residues A10-U12 from MD simulations in the presence and absence of Mg^{2+}. **a/b** Time evolution of the pseudorotation (phase) angle for the hairpin residues A10, U11 and U12 in the RNA-RNA (blue) or RNA-DNA contact (orange) in the presence and absence of Mg^{2+}. **c/d** Autocorrelation of the pseudorotation angle.

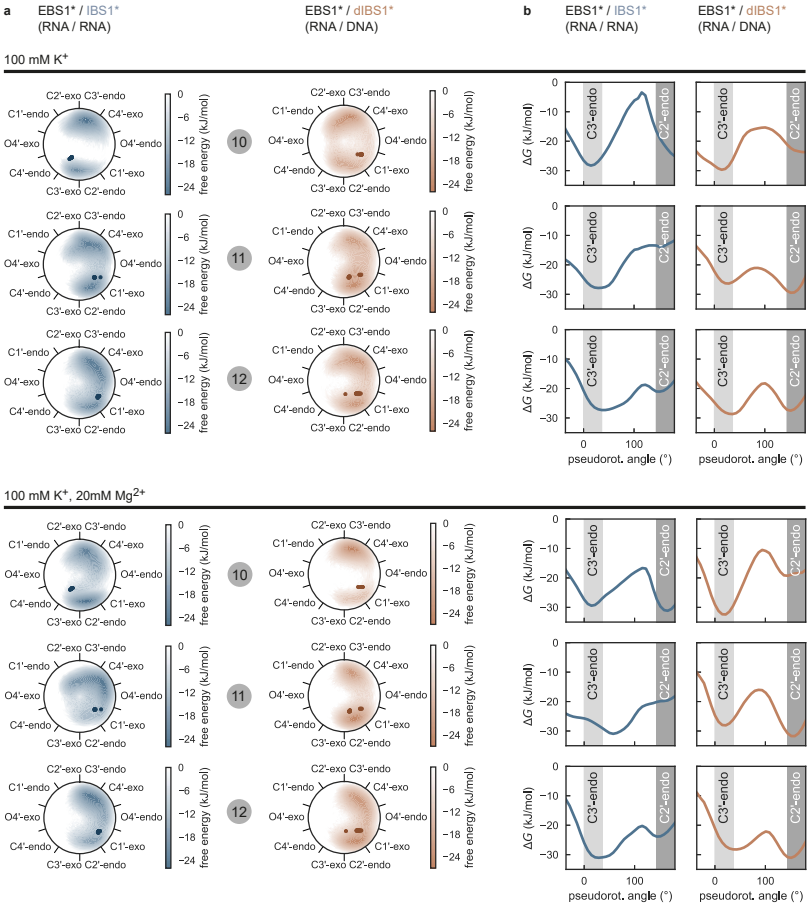


Fig. 4.27 | Sugar pucker conformations of EBS1* residues A10-U12 from MD simulations in the presence and absence of Mg^{2+}. **a** Pseudorotation cycles for A10, U11 and U12. Distributions of pucker phase and amplitude are color coded by their free energy. The puckers of the 18 (EBS1*/IBS1*, blue) or 20 (EBS1*/dIBS1*, orange) lowest energy NMR structures are indicated by dots. **b** Free energy profiles of the pseudorotation angle (pucker phase) for all residues of IBS or dIBS.

Chapter 5

FRET-assisted modeling of riboswitch structural dynamics

Fabio D. Steffen, Sofia Gallo, Roland K. O. Sigel[‡], Richard Börner[‡]

Department of Chemistry, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

Abstract: Cobalamin riboswitches regulate vitamin B_{12} homeostasis by structural rearrangements in response to metabolite binding. Here, we introduce a molecular modeling strategy complemented by single-molecule FRET to probe the dynamics of the *btuB* riboswitch from *Escherichia coli*. We find that the ribosomal binding site hairpin, a key element in gene control, samples open and closed conformations in the absence of adenosylcobalamin. Mg^{2+} modulates this equilibrium and prepares the aptamer for interaction with the metabolite. Folding of the RBS stem coincides with a global collapse of the RNA, suggesting a signal relay from the aptamer via the pseudoknot to the distal parts of the expression platform. The presented modular workflow integrating FRET into *de novo* prediction will likely accelerate the elucidation of other transient conformational RNA ensembles where prior structural information is sparse or incomplete.

manuscript in preparation **2020**

5.1 Introduction

Integrative structural modeling of nucleic acids and proteins has surged in recent years owing to an increasing availability of experimental restraints obtained by various of biophysical techniques.[221] Such restraints include electron density maps[162,340], distances from FRET[204,220,222] or electron paramagnetic resonance (EPR)[341], and chemical cross-links detected by quantitative mass-spectrometry.[223,224] FRET combined with single-molecule detection lends itself to hybrid modeling as it can resolve short-lived conformational states intractable by structural biology techniques having a high spatial but insufficient temporal resolution.[129,177,204,222] The potential of single-molecule FRET unfolds in conjunction with a (near)-atomic representation of a macromolecule in a particular state. An initial structure or structural ensemble could be obtained from X-ray crystallography, NMR, cryo-EM or *de novo* prediction. During the modeling process, computed decoys are scored and filtered based on their consistency with a minimally redundant network of donor-acceptor dye pairs.[204,220,222]. The experimentally measured FRET efficiencies are linked to the candidate structures via a dye model that describes the constrained movements of the attached fluorophore.[171,188,199,204,205,258,342] Once validated against experiments and simulations, these probabilistic search algorithms, inspired by the global positioning system (GPS), can replace a computationally more expensive all-atom treatment of the dye.[171,201–203,343]

Riboswitches are conformationally dynamic RNAs and therefore an ideal target for integrative modeling. They selectively sense metabolites including coenzymes[33,38,54], signaling molecules[44], amino acids[90,256] and ions[47,53]. One of the first validated riboswitches was found in the 5′-untranslated region of the *btuB* gene in *Escherichia coli* and *Salmonella typhimurium* (*Sth*).[38,62] It controls the expression of an outer-membrane transporter for cobalamins and is thus integral to the metabolism of vitamin B_{12}.[60,61]. Cobalamin aptamers are amongst the most widespread riboswitches in bacteria.[39,66] Regulation of their associated genes occurs either co-transcriptionally through formation of a terminator hairpin[65], or alternatively by restricting access to the ribosomal binding site and thus preventing translation initiation.[61,62] The *E. coli btuB* is thought to be translationally controlled, however, the exact mechanism of signal transmission from the aptamer core to the

expression platform remains unclear. Specifically, the dynamics through which ligand binding is communicated to the distal parts of the RNA, where most of the structural rearrangements take place, has not yet been addressed in detail.

Two different homologs of the *E. coli btuB* leader have been crystallized in their metabolite bound state.[69,86] They reveal an intricate tertiary architecture consisting of two coaxially stacked helices connected via a four-way junction and peripheral elements docking into the core to reinforce the binding site. Adenosylcobalamin (AdoCbl) contacts the RNA almost exclusively by van der Waals interactions. In close proximity of the B_{12} pocket is a pseudoknot which presumably orients the expression platform relative to the aptamer and is therefore assumed to be involved in transducing the metabolite binding signal to the place where the genetic decision is made.[87] For a translationally controlled riboswitch this is the Shine-Dalgarno sequence. The hairpin that grants access to the ribosomal binding site (RBS) is missing in both crystal structures though.[86,87] This prompted us to build a structural model of the full-length *btuB* riboswitch from *E. coli* including the RBS and residues up to the the start codon. For this purpose, we take conserved features of the homologous crystal structures and predict the missing peripheral motifs *de novo*. With a refined dye model, single-molecule FRET restraints are then mapped onto the structure to filter the computed models. We find that the RBS stem loop hairpin is conformationally dynamic in line with its role as a gene-regulatory element.

5.2 Results

5.2.1 The core of the E. coli aptamer is evolutionary conserved

We started our modeling pipeline by aligning the *E. coli* sequence of the *btuB* riboswitch to a consensus model of cobalamin riboswitches from the Rfam database (Fig. 5.1a and Supplementary Fig. 5.4). The consensus alignment contains the aptamer from the beginning to the end of helix P1 but does not include the structurally less conserved expression platform. The multi-sequence alignment and the consensus secondary structure agreed with previous chemical footprinting assays of the *E. coli* riboswitch.[66,245,344,345] We additionally annotated tertiary contacts in the secondary structure diagrams by comparing conserved or covary-

ing nucleotides between the *E. coli* and the crystallized *T. tengcongensis* (*Tte*)[86] and *S. thermophilum* (*Sth*)[87] riboswitches. The aligned *E. coli* sequence was then threaded onto the *Tte* crystal structure to give a homology model of the aptamer core (Fig. 5.1b). The missing peripheral domains were built *de novo* in three separate steps using Rosetta's fragment assembly of RNA with full atom refinement (FARFAR2).[156]
We decided against modeling the entire structure at once for two reasons: first, it was evident from the template structures that domains P8-P12 and P13/RBS are pointing in opposite directions and are thus largely independent. Modeling them separately and ligating the best models *a posteriori* allowed to increase the sampling of either region. Modeled in a single run, they would exceed the threshold of about 80 nucleotides beyond which *de novo* prediction becomes increasingly challenging as determined in a recent, comprehensive benchmark.[156] Wherever loops or helices are expected to sterically affect each other, we modeled them together. Secondly, the pseudoknot along with the parallel coaxial stack P1-P3-P6 is anticipated to critically determine the orientation of the emanating stem P13. According to the same rational as above, we kept the pseudoknot motif as small as possible but as large as sterically needed to prevent clashing in subsequent modeling steps.
Hence, we began by reconstituting small segments inside the aptamer which were left out in the homology model due to variations between the query and the template. These include the loop regions of P2, P11, a stretch of two nucleotides in P4 and most notably the kissing loops L5 and L13.

In both crystal structures P5 makes a sharp kink at the proximal stem-to-loop transition in order to accommodate the complementary L13. This kink is reproduced in the fragments (Fig. 5.1c). L13 is extended by two nucleotides compared to the *Tte* which appears to make the loop at the distal end of the pseudoknot more flexible (Supplementary Fig. 5.5b). While not directly evident from the crystal structure due to a mutation which was introduced to improve crystallization, the pseudoknot of the wild type *Tte* sequence is destabilized by a C-A mismatch. The mismatch was suggested to facilitate the reversibility of pseudoknot formation, thereby obviating a constitutively repressed riboswitch.[86] The observed conformational flexibility in the *E. coli* loop might serve a similar purpose and favor the dissociation of the GC-rich and perfectly WC base-paired kissing loops (A47-G51 / C187-U191).[346] While modeling the pseudoknot in this first phase, we included stem nucleotides up to

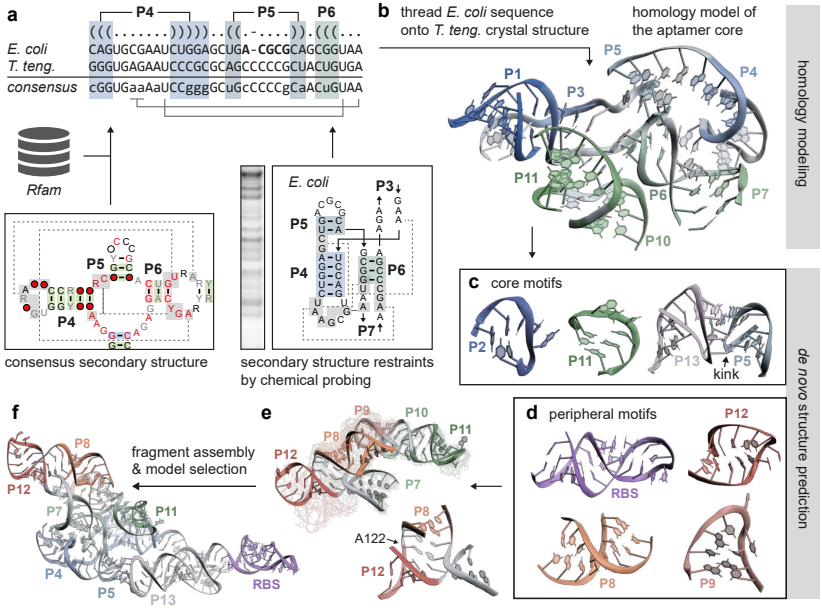


Fig. 5.1 | **Homology and de novo modeling of the E.coli btuB riboswitch with Rosetta.**
a The *E. coli btuB* sequence is aligned to the cobalamin consensus model from the Rfam database. **b** A homology model is built by replacing residues from *T. tengcongensis* in the crystal structure with those from *E. coli*. **c/d** Core and peripheral motifs are predicted *de novo* and grafted onto the homology model. **e** Overlay of the top-five lowest energy structures shows that the variability is largest in P12. Adenine 122 at the base of the three-way junction was previously postulated to be a Mg^{2+} binding site.[345] Coordination of divalent metal ion at this position would help to structure P12. **f** Fully assembled, low-energy conformation of the *E.coli btuB* riboswitch with regions P1-P13/RBS highlighted in color.

the base pair U180/A202, corresponding to the end point of the original *btuB* riboswitch construct (denoted thereafter as $btuB_{1\text{-}202}$).[38,66,344]

5.2.2 The RBS hairpin and P12 are structurally dynamic

Next, we set out to build the peripheral domains of the riboswitch (Fig. 5.1d). On the one hand, these are the helical stems P8 to P12 describing a U-shaped turn which docks back into the core via two A-minor interactions.[87] The stems P7 and P10/11 are included as seeds to position the modeled domains with respect to the aptamer. The models recapitulate the topology seen in the crystal structures and additionally include two three-way junctions which are either unresolved or absent in the other bacterial sequences. The base of P12 was previously identified as a coordination site for Mg^{2+}.[345] We expect that binding of a divalent ion would stabilize this motif akin to what has been observed in the κ three-way junction of group II introns.[347]
The third and final step involves the assembly of the ribosomal binding site hairpin and the tethering of expression platform and aptamer via helix P1. Inspection of the best-scoring models revealed significant dynamics of the hairpin (Supplementary Fig. 5.5c). In one cluster, the RBS stem is stacked on P13. This coaxial alignment is initiated by P4 and reaches from the aptamer into the expression platform, bridged by the kissing loops L5 and L13. Other low-energy models feature an RBS hairpin swinging towards the exposed P11 stem loop. Interaction of the RBS with P11 is not indicated by any secondary structure as it involves non-WC interactions. The proximity of the two motifs is only evident in three dimensions. Docking of the RBS hairpin on P11 is probably transient and rather weak and may therefore not necessarily be apparent in footprinting assays. It does represent a interesting target for single-molecule FRET though, as discussed further below.

5.2.3 Multiple accessible-contact volumes filter Rosetta models

Given the range of conformations recovered for the RBS hairpin, we asked whether a set of FRET measurements to distant parts of the riboswitch would support the orientation of the RBS stem-loop within a coaxially stacked helical bundle or rather one of the kinked variants. Previously, we labeled the riboswitch internally on adenines A35 and A213 as well as on the 3′-terminus (see also Chapter 3).[245,314] To calculate

FRET efficiencies for the two distance vectors we computed accessible-contact volumes, which estimate the sterical constraints imposed on the fluorophores by the RNA. Before applying the ACV model on the riboswitch, we benchmarked it against molecular dynamics simulations with fluorophores explicitly included in the force field. For these tests we used different single and double-stranded model systems (Supplementary Figs. 5.6). We have introduced the contact volume[171] as an extension of the original AV approach[188,258]. It reweights the geometrical point cloud by the residual anisotropy and in this way accounts for a fraction of dyes interacting with the biomolecule.[171,204,205] Such fluorophore stacking has been found predominantly at the termini of helical elements and on single-stranded overhangs.[167,171,259] The effect on the measured FRET efficiency is particularly pronounced if the ACVs are aligned coaxially, hence if they are either pointing towards or away from each other. This is demonstrated by a DNA hairpin with fluorophores attached to opposite sides of the helix (Supplementary Fig. 5.6b/c). The AV alone is shown to slightly underestimate the measured FRET efficiency. Upon reweighting with the contact volume the mean inter-dye distance is shortened and the FRET efficiency increases. Since folded RNA and proteins adopt a more spherical shape than elongated DNA helices, the CV is expected to be particularly relevant for biomolecules, as demonstrated recently for the proteins atlastin and lysozyme.[204,222] Next, we calculated mean FRET efficiencies between the 3′-end and the two adenines for the 200 lowest energy structures of the *btuB* riboswitch (Fig. 5.2a/b and Supplementary Fig. 5.8). By confocal spectroscopy we measured FRET histograms for the same fluorophore positions in the presence of 20 mM $MgCl_2$ and 50 mM KCl without AdoCbl (Fig. 5.2c/d). Overall, the computed and experimental mean efficiencies of the best models agree fairly well and the inclusion of the contact volume indeed shifts the predicted efficiency closer towards the experimentally observed values. Yet, the FRET values are not in a very sensitive range for precise distance measurements. Moreover, two FRET constraints alone are insufficient to pinpoint the location of the RBS with respect to the aptamer. In particular, they cannot monitor a potential docking reaction on P11. Introducing a donor label near the P11 loop in future experiments could help to discriminate between the best scoring models.

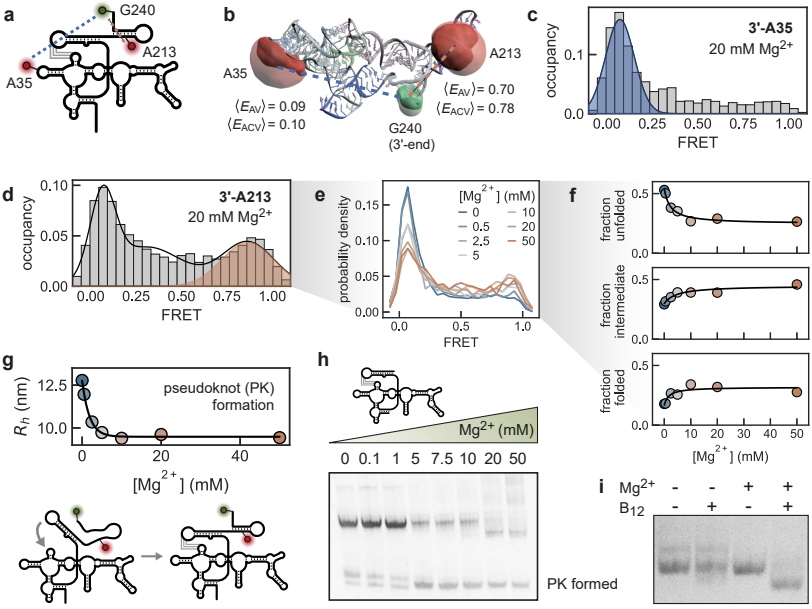


Fig. 5.2 | **Single-molecule FRET probes riboswitch tertiary structure and dynamics.**
a Secondary structure drawing of the full length *btuB* riboswitch of *E.coli* with labeled residues indicated. **b** Accessible contact volumes and FRET predictions for the distance coordinate 3′-A35 and 3′-A213 Weights of the contact volume were determined from the residual anisotropies of the dyes[245] (3′-end: $r_\infty = 0.35$, $w_{CV} = 0.92$; A35: $r_\infty = 0.22$, $w_{CV} = 0.58$; A213 $r_\infty = 0.24$, $w_{CV} = 0.63$) **c/d** Experimental FRET histogram recorded in the presence of 20 mM $MgCl_2$ and 50 mM KCl at pH 7.5. The FRET population corresponding to the modeled structure is shaded in blue. **e/f** Mg^{2+} dependent formation of the RBS hairpin in the context of the riboswitch. Increasing intermediate and high FRET efficiencies indicate folding of the hairpin. **g** A concomitant decrease of the hydrodynamic radius obtained from FCS suggests that formation of the pseudoknot (PK) seeds a coaxial stack which is extended by the RBS hairpin. **h** Native polyacrylamide gel electrophoresis (PAGE) showing Mg^{2+} dependent collapse of the RNA for the shortened $btuB_{1\text{-}202}$.[344] **i** AdoCbl-induced compaction of the pre-folded RNA visualized by native PAGE.

5.2.4 The RBS hairpin equilibrates open and closed conformations

The distance coordinate between the 3′-terminus and A213 is of interest, less for FRET mapping than for probing the accessibility of the Shine-Dalgarno sequence. The RBS stem is the entry point for the ribosome to bind the RNA and translate the *btuB* gene.[62] In the presence of AdoCbl, the RBS hairpin is formed and gene expression is turned down. We found that the isolated hairpin is thermodynamically stable also in the absence of AdoCbl and divalent ions, owing to its high GC-content (Supplementary Fig. 5.9). In these FRET measurements, A213 was substituted for a uracil, which is labeled with Cy5, such that the dye attachment (at C5) does not interfere with the Watson-Crick (WC) face. Conversely, in the full riboswitch attachment of the fluorophore via an ethenoadenosine blocks the WC edge and the nucleotide will have to rearrange to present its Hoogsteen or sugar edge for interaction (Supplementary Fig. 5.9i/j). Although the duplex is destabilized it is still able to form, yet with a noticeable Mg^{2+} dependence (Fig. 5.2e/f and Supplementary Fig. 5.10). Increasing levels of Mg^{2+} depopulate the low in favor of the intermediate and high FRET states corresponding to a base-paired hairpin. Mg^{2+}-induced half-maximal saturation is reached around 3 mM of divalent ions which coincides with a compaction of the RNA as seen by a decrease in hydrodynamic radius via FCS (Fig. 5.2g). The correlated global collapse and local stem formation suggests a possible coupling between the pseudoknot and the RBS hairpin and thus between aptamer and expression platform. The driver is probably the aptamer, as compaction is induced at similar Mg^{2+} concentrations also in the absence of the RBS hairpin ($btuB_{1\text{-}202}$, Fig. 5.2h).[344] Addition of AdoCbl consolidates the fold as observed on a native polyacrylamide gel (Fig. 5.2i and Supplementary Fig. 5.11).

5.3 Discussion

The *S. typhimurium* cobalamin riboswitch was one of the earliest RNA puzzles[149] and was recently revisited as part of a benchmark for FARFAR2.[156] Its intricate tertiary architecture has posed a significant challenge for computational modeling. Some key interactions between the periphery and the core were missed since non-WC base pairs are inherently difficult to predict.[149,150] Here, we gathered biochemical and structural information from nucleotide covariance and crystallography

to build a model of the full-length *btuB* riboswitch of *E. coli*. We implemented all-atom and coarse-grained descriptions of the fluorophores to predict FRET efficiencies *in silico* and relate them to experiments. These probabilistic dye-linker models are particularly useful for evaluating possible distance coordinates and their sensitivity, given that a structure is available. As such they help in the rational design of FRET experiments.
For the *btuB* riboswitch we chose a local FRET coordinate that monitors the accessibility of the ribosomal binding site hairpin. Our single-molecule measurements showed that the RBS hairpin exists in open and closed forms in the absence of AdoCbl. Magnesium(II) compacts the RNA by knotting together two loops, L5 and L13, and further promotes closing of the RBS hairpin. AdoCbl resides at a linchpin where aptamer and expression platform are joined together.[86,87] Beside P1 and the pseudoknot, a potential interaction of the RBS with P11 could establish a third contact point between the metabolite sensing domain and the downstream regulatory unit. This interaction is probably weak but could contribute to the compaction of the riboswitch.

We propose a conformational selection mechanism for the *btuB* riboswitch in which the RNA samples multiple conformations of tight and relaxed helical stacks involving regions around the pseudoknot and the RBS hairpin (Fig. 5.3). Their prevalence is modulated by the availability of divalent ions. AdoCbl then engages with a pre-folded structure, consolidates this conformer and thereby reaches a genetic decision. Such a mechanism bears similarity with allostery in proteins[348,349] and has been proposed for other riboswitches as well.[80,101,350] In the cell, transcription additionally adds a time-constraint on the equilibrium of metabolite and riboswitch.[67,71] To ensure sufficient time for AdoCbl to interact with the folding aptamer, the polymerase was found to pause at different sites while transcribing the expression platform.[76] Strategic pausing prevents an incompletely assembled aptamer to base-pair with complementary parts of the emerging expression platform (conformer i in Fig. 5.3). If the riboswitch commits to this kinetically trapped anti-aptamer structure, access to the ribosomal binding site is granted and gene expression is turned on. The low FRET peak in our post-transcriptional measurements in the absence of AdoCbl could indeed arise from such an alternative fold of aptamer and expression platform. While RNA-only conformers have mostly resisted structure determination due to poor crystal packing, Rosetta was recently shown to resolve

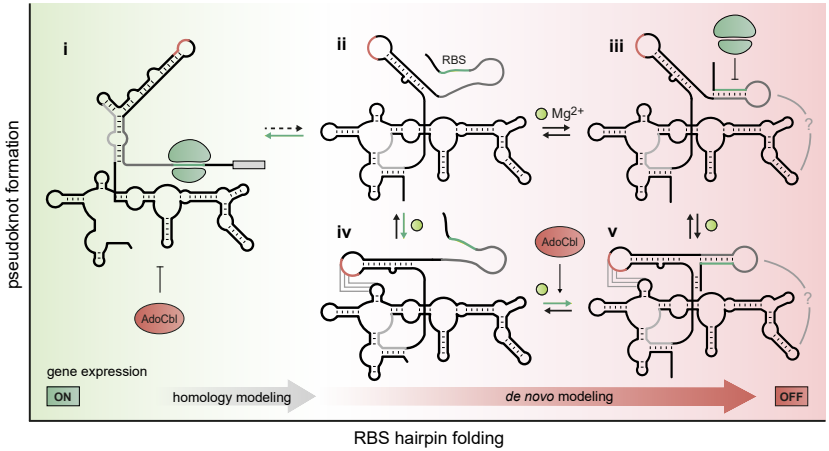


Fig. 5.3 | **Extract of parallel folding pathways of the btuB riboswitch.** Depicted are five possible conformations of the aptamer and expression platform probed by single-molecule FRET and modeled computationally. Green and red shaded areas indicate structures where gene expression is presumed to be activated or repressed. Conformation **i** is a kinetically trapped variant in which the sequence upstream of the RBS base-pairs with the aptamer, allowing the ribosome to interact with the accessible RBS. This conformer could contribute to the low FRET state explored by the FRET coordinate 3′-A213. In surface immobilized experiments this state would probably appears as static, low FRET time traces. Species **ii** and **iii** are riboswitches which have not yet undergone compaction by forming the pseudoknot. In **iii** the RBS is sequestered by the hairpin and the ribosome cannot bind. In conformations **iv** and **v** the pseudoknot has formed upon addition of Mg^{2+}. The Mg^{2+} dependence of the equilibria **ii-iii** and **iv-v** was probed by single-molecule FRET. Adenosylcobalamin is suggested to consolidate the folded species **v**. This may be aided by a yet unconfirmed interaction between P11 and the RBS hairpin. We speculate the main flux to go along the pathway indicated by the green arrows. However, single-molecule kinetic data is required to test this hypothesis. The final *in silico* predicted structures correspond to the gene repressed conformation **v** with AdoCbl bound although the metabolite was not explicitly included in the modeling. The need for RNA structure prediction algorithms like Rosetta to include small-molecules has been highlighted by the RNA puzzle trials.[149,150]

global RNA-only folds by modeling coordinates into cryo-EM maps.[162] FRET-assisted and density-guided structure prediction are largely complementary and thus present interesting new perspectives for integrative modeling of such RNA-only conformations.

5.4 Material and Methods

5.4.1 RNA transcription and purification

The btuB riboswitch construct from *E. coli* used herein contains the full length aptamer and expression platform up to the start codon. At the 5′-end of the aptamer a 35 nucleotide extension sequence was introduced to allow surface immobilization via a complementary DNA oligonucleotide.[245] The RNA was transcribed *in vitro* from linearized plasmid DNA using T7 RNA polymerase.[351]. The transcribed RNA was purified by 10 % denaturing PAGE, electroeluted and recovered by precipitation in EtOH. Finally, the RNA was desalted over a Vivaspin (MWCO: 5000 g mol^{-1}).

5.4.2 RNA fluorescence labeling

The purified riboswitch was labeled covalently at the 3′-end and on adenines A35 and A213 using a DNA-guided functionality transfer reaction and click chemistry as described previously[245,352] and recently reviewed[314] (see Chapter 3). In brief, DNA probes were designed to hybridize near the target site to break up the local base-pairing. Another antisense DNA oligo was annealed directly up- or downstream of the labeling position. This oligo carries a propargyl group and a diol moiety. Upon oxidation with $NaIO_4$ it reacts with the exocyclic amine of the adenine to transfer the alkyne onto the RNA. The modified RNA was then treated with an azide functionalized Cy5 dye. Concurrently, $NaIO_4$ oxidizes the ribose at the 3′-end which was subsequently reacted with a Cy3-hydrazide to yield the double-labeled RNA.

5.4.3 Native gel electrophoresis

The riboswitch was refolded at pH 7.5 in 34 mM Tris-HCl, 66 mM HEPES and 50 mM KCl. For this purpose the RNA was heated to 70 °C for 5 min in the presence of Cy3 and Cy5 labeled complementary DNA oligonucleotides (Supplementary Table 5.1). After cooling down to room temperature 0.1 mM–50 mM $MgCl_2$ and, if indicated, 100 µM AdoCbl was added under exclusion of light followed by 30 min incubation at 37 °C and room temperature respectively. Samples were then loaded on a 6% (w/v) polyacrylamide gel and run at 4 °C in the dark. Gels were visualized under UV (254 nm) and by recording the fluorescence emission

on a Typhoon FLA 9500 scanner (GE Healthcare) with filters BPG1 (570DF20) for Cy3 and LPR (665LP) for Cy5. Bands were quantified with ImageQuant 8.1 (GE Healthcare).

5.4.4 Confocal FRET spectroscopy

Free-diffusion single-molecule experiments were carried out on a commercial MicroTime 100 confocal microscope (PicoQuant) with pulsed interleaved excitation (PIE)[353] at 518 nm (LDH-D-C-520, PicoQuant) and 638 nm (LDH-D-C-640, PicoQuant) respectively and a repetition rate of 40 MHz. The instrument was equipped with an Olympus UPlanSApo60×1.20-W water-immersion objective and TimeHarp 260 counting electronics (PicoQuant). Photons were spectrally separated by a T635lpxr dichroic mirror (Chroma), filtered with a FF01-582/64 (Semrock) and a H690/70m (Chroma) bandpass and collected on a SPCM-AQRH single-photon avalanche detector (Excelitas). Data was analyzed in the Matlab based software package PAM.[354] Single-molecule bursts were identified with a sliding time window search[355] and photoblinking events were removed with the ALEX-2CDE filter.[189] To eliminate contributions from donor and acceptor-only molecules, an intensity threshold for red photons $F^{\mathrm{A_{em}}}_{\mathrm{A_{ex}}} > 20-50$ photons and a stoichiometry filter $S > 0.1-0.2$ were applied. The FRET efficiency E was calculated as the ratio of corrected donor and acceptor intensities

$$E = \frac{\left(F^{\mathrm{A_{em}}}_{\mathrm{D_{ex}}} - \alpha F^{\mathrm{D_{em}}}_{\mathrm{D_{ex}}} - \delta F^{\mathrm{A_{em}}}_{\mathrm{A_{ex}}}\right)}{\gamma F^{\mathrm{D_{em}}}_{\mathrm{D_{ex}}} + \left(F^{\mathrm{A_{em}}}_{\mathrm{D_{ex}}} - \alpha F^{\mathrm{D_{em}}}_{\mathrm{D_{ex}}} - \delta F^{\mathrm{A_{em}}}_{\mathrm{A_{ex}}}\right)} = \frac{F'^{\mathrm{A_{em}}}_{\mathrm{D_{ex}}}}{\gamma F^{\mathrm{D_{em}}}_{\mathrm{D_{ex}}} + F'^{\mathrm{A_{em}}}_{\mathrm{D_{ex}}}}. \quad (5.1)$$

Here, α corrects for donor leakage into the acceptor channel and is calculated from a donor-only species as $\alpha = \frac{F^{\mathrm{A_{em}}}_{\mathrm{D_{ex}}}}{F^{\mathrm{D_{em}}}_{\mathrm{D_{ex}}}}$. Similarly, δ corrects for direct excitation of the acceptor at the donor excitation wavelength. It is calculated from an acceptor-only species as $\delta = \frac{F^{\mathrm{A_{em}}}_{\mathrm{D_{ex}}}}{F^{\mathrm{A_{em}}}_{\mathrm{A_{ex}}}}$. Finally, γ accounts for differences in quantum yield ϕ of the donor and acceptor dye and their relative detection efficiencies η, giving $\gamma = \frac{\phi_{\mathrm{A}}\,\eta_{\mathrm{A}}}{\phi_{\mathrm{D}}\,\eta_{\mathrm{D}}}$.[190,191,207] The FRET-independent stoichiometry is defined as

$$S = (1+\beta)^{-1} \quad (5.2)$$

where β stands for the ratio of excitation powers $I_{\mathrm{A_{ex}}}/I_{\mathrm{D_{ex}}}$ and dye absorption cross-sections σ, giving $\beta = \frac{I_{\mathrm{A_{ex}}}\sigma^{\mathrm{A}}_{\mathrm{A_{ex}}}}{I_{\mathrm{D_{ex}}}\sigma^{\mathrm{D}}_{\mathrm{D_{ex}}}}$.[190,191] In practice, γ

and β were obtained from the slope and intercept upon plotting the inverse stoichiometry $1/S$ against the cross-talk corrected proximity ratio E_{PR}[190]

$$1/S = 1 + \gamma\beta + \beta(1-\gamma)E_{\mathrm{PR}} = \Omega + \Sigma E_{\mathrm{PR}} \tag{5.3}$$

with $\Omega = 1+\gamma\beta$ and $\Sigma = \beta(1-\gamma)E_{\mathrm{PR}}$ which yields after rearrangement

$$\beta = \Omega + \Sigma - 1 \quad \text{and} \quad \gamma = \frac{\Omega - 1}{\Omega + \Sigma - 1}. \tag{5.4}$$

The resulting FRET histograms were fitted to a Gaussian mixture model with up to three components.

Diffusion coefficients D were determined by fitting the auto- and cross-correlation curves from FCS with a diffusion term and a triplet component

$$g(\tau) = \frac{1}{\langle N \rangle}\Big(1 + \frac{T}{1-T}e^{-t/\tau_T}\Big)\Big(\frac{1}{1+\tau/\tau_D}\Big)\Big(\frac{1}{1+(\omega_r/\omega_z)^2\tau/\tau_D}\Big)^{1/2} \tag{5.5}$$

$$\text{with} \quad \tau_D = \frac{\omega_r^2}{4D} \tag{5.6}$$

where $\langle N \rangle$ is the average number of molecules in the focus, T is the triplet component, τ_T and τ_D are the triplet and translational diffusion times, and ω_r and ω_z the lateral and axial dimensions of the confocal volume. The hydrodynamic radius R_{h} was calculated from the diffusion coefficient using the Stokes-Einstein equation

$$R_{\mathrm{h}} = \frac{k_{\mathrm{B}}T}{6\pi\eta D} \tag{5.7}$$

where k_{B} is the Boltzman constant, T the temperature and η the viscosity of the medium.

5.4.5 Ensemble TCSPC

Lifetime and anisotropy decays were recorded on a Fluorolog FL3-222 (Horiba) spectrophotometer equipped with a double-grating emission monochromator and a PPD-900 photon counting detector. Samples were excited at 504 nm (DD-510L, Horiba) and 633 nm (DD-635L,

Horiba) with a repetition rate of 20 MHz. Decays were fitted to a series of exponentials by iterative reconvolution with the instrument response function (IRF) using the custom written Python package LifeFit (https://github.com/fdsteffen/
Lifefit). Anisotropy decays were calculated from the polarization resolved decays

$$r(t) = \frac{I_{\parallel}(t) - G\, I_{\perp}(t)}{I_{\parallel}(t) + 2G\, I_{\perp}(t)} \tag{5.8}$$

and analyzed with a local-global rotation model[166,171]

$$r(t) = \left[(r_0 - r_\infty)\, e^{-t/\tau_{r,loc}} + r_\infty\right] e^{-t/\tau_{r,glob}}. \tag{5.9}$$

5.4.6 Homology and de novo modeling with Rosetta

Nucleotide sequences of the *T. tengcongensis* (PDB ID: 4gma), *S. thermophilum* (PDB IDs: 4gma) and *E. coli* B_{12} aptamers were aligned to the consensus sequence of cobalamin riboswitches from the Rfam database[356] (ID: RF00174[65,66,357,358]) using Infernal 1.1.3.[359] The primary sequence of the *E. coli* btuB aptamer was threaded onto the 3D structure of the *Tte* core using Rosetta's `rna_thread` application.[161,360] The stem loop P13, which is part of the pseudoknot, was aligned manually to the *Tte* expression platform. P13 and the ribosomal binding site (RBS) hairpin were then modeled *de novo* using FARFAR2[156] which is part of Rosetta 3.12. Specifically, the FARFAR2 protocol implements a base-pair step sampling algorithm that retains the WC geometry of the bases but allows the backbone to move. In a recent benchmark, this newly implemented set of Monte-Carlo moves outperformed previous protocols that were using ideal, fixed helices as inputs.[156] Moreover, the `rna_denovo` executable, which is the heart of FARFAR2, draws on an updated fragment library (*RNA18_HUB_2.154_2.5.torsions* and *RNA18_HUB_2.154_2.5.jump*)[156,361] and features a new scoring function *rna_res_level_energy4.wts* ("res4")[163] for full-atom refinement.
Models were generated with the following command (here given exemplarily for the first modeling step, i.e. assembly of the aptamer core without P8/9/12 and the RBS hairpin) run in parallel on multiple cores

```
rna_denovo -fasta aptamer.fasta -nstruct 1000 \
-secstruct_file aptamer.txt -minimize_rna true \
-cycles 20000 -s 4gma_homology_template.pdb
```

with the fasta sequence and secondary structure (in dot-bracket notation) provided as separate files

```
>aptamer A:1-63 A:88-111 A:139-161 A:180-202
gccgguccugugaguuaauagggaauccagugcgaaucuggagcug
acgcgcagcgguaagga,gacacugccauucgguggggaaguc,ucc
aagcccgaagaccugccggc,ucaucaucgcguaauauugauga


((((((.((...)).....(((...(((((.......)))))..((
[[[[[)).(((...(((,(((.(((((....)))))...))),)))
....))).....))))))))),((((((.]]]]].....))))))
```

5.4.7 Multiple accessible-contact volume (mACV)

Accessible-contact volume simulation employ a dye model where the fluorophore is approximated by three radii $R_{\mathrm{dye},1-3}$ and a flexible linker with a length $L_{\mathrm{linker,max}}$ and width $W_{\mathrm{linker,max}}$.[258,362] Dimensions of the individual components (base, linker, dye) are summarized in Supplementary Fig. 5.12. Accessible dye positions are searched on a grid surrounding the biomolecule and evaluated with respect to the maximum linker length using a shortest path (Dijkstra) algorithm. Contact volumes of a given thickness W_{CV} (usually set equal to one of the dye radii) are reweighted by the experimentally determined residual anisotropy of the fluorophores.[171,204] The weight of the contact volume is given by

$$w_{\mathrm{CV}} = 1 - \frac{r_0 - r_\infty}{r_0} \tag{5.10}$$

where r_0 and r_∞ are the fundamental and residual anisotropy respectively. The ACV algorithm is available as a Python module to calculate accessible volumes in high-throughput (using a series of PDBs or a molecular dynamics trajectory). Alternatively, ACVs can be computed for models in a single PDB file using a PyMOL plugin that grants visual feedback (https://github.com/fdsteffen/fretraj). This is particularly useful for designing new FRET experiments. Both implementations build upon the LabelLib library (https://github.com/FluorescenceTools/LabelLib.git) but also feature a Python implementation of the same algorithm.

5.4.8 FRET prediction

Within the ACV framework, the mean donor acceptor distance $\langle R_{\mathrm{DA}} \rangle$ is calculated from a discrete set of distances between randomly sampled points in the two ACVs

$$\langle R_{\mathrm{DA}} \rangle = \frac{1}{n\,m} \sum_{i=1}^{n} \sum_{j=1}^{m} \|\mathbf{R}_{\mathrm{A},i} - \mathbf{R}_{\mathrm{D},i}\| \tag{5.11}$$

where $\mathbf{R}_{\mathrm{D}}$ and $\mathbf{R}_{\mathrm{A}}$ are the position vectors of the points i and j in the donor and acceptor cloud respectively.[191,258]
Similarly, the mean FRET efficiency $\langle E \rangle$ is calculated from the same set of donor-acceptor distances and the Förster radius R_0

$$\langle E \rangle = \frac{1}{n\,m} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{1 + \|\mathbf{R}_{\mathrm{A},i} - \mathbf{R}_{\mathrm{D},i}\|^6 / R_0^6} \tag{5.12}$$

The FRET averaged distance $\langle R_{\mathrm{DA}} \rangle_E$ is then computed from the mean FRET efficiency

$$\langle R_{\mathrm{DA}} \rangle_E = R_0 (\langle E \rangle^{-1} - 1)^{1/6} \tag{5.13}$$

Finally, the distance between the mean dye position in each ACV is calculated from the mean of the position vectors $\mathbf{R}_{\mathrm{D},i}$ and $\mathbf{R}_{\mathrm{A},j}$

$$R_{\mathrm{MP}} = \|\langle \mathbf{R}_{\mathrm{A},j} \rangle - \langle \mathbf{R}_{\mathrm{D},i} \rangle\| \tag{5.14}$$

$$= \Big\| \frac{1}{m} \sum_{j=1}^{m} \mathbf{R}_{\mathrm{A},j} - \frac{1}{n} \sum_{i=1}^{n} \mathbf{R}_{\mathrm{D},i} \Big\| \tag{5.15}$$

In MD simulations with explicit fluorophores, FRET efficiencies are computed from photon bursts which are averaged over a single or multiple trajectories as described previously.[201,343] Burst are sampled from a Markov-chain characterized by the photon emission rates $k_{\mathrm{D,f}}$ and $k_{\mathrm{A,f}}$, the internal conversion rates $k_{\mathrm{D,i}}$ and $k_{\mathrm{A,i}}$ and the time-dependent transfer rate $k_{\mathrm{T}}(t)$. The transfer rate is updated for each time-step of the Monte-Carlo simulation by the change in inter-dye distance R_{DA} and the orientation factor κ^2 of the underlying MD trajectory

$$k_{\mathrm{T}}(t) = (k_{\mathrm{D}} + k_{\mathrm{D,i}}) \Big(\frac{R_0(t)}{R_{\mathrm{DA}}(t)} \Big)^6 \tag{5.16}$$

with

$$k_{\mathrm{D}} = \frac{\phi_D}{\tau_D}, \quad k_{\mathrm{D,i}} = \frac{1-\phi_D}{\tau_D} \tag{5.17}$$

and

$$R_0(t) = R_{\mathrm{const}}\kappa^2(t) = \frac{54\,\text{Å}}{\sqrt[6]{2/3}}\kappa^2 \quad \text{(for Cy3-Cy5)} \tag{5.18}$$

where ϕ_D and τ_D are the fluorescence quantum yield and lifetime of the donor dye.[201,343]

5.4.9 Fluorophore and linker parameters for in silico labeling

Bonded and non-bonded parameters for the fluorophores are taken from the AMBER-DYES package.[363] Parameters for the linker atoms were derived by analogy with similar terms in the AMBER force field using the Antechamber Python Parser interface (Acpype).[364] Linker geometries were optimized with GAUSSIAN 09[365] first at the B3LYP/6-31G* level of theory and followed by electrostatic potential (ESP) calculations with Hartree-Fock. Partial charges were calculated by two-stage restrained electrostatic potential (RESP)[366] fitting, implemented in Antechamber.[367] For this purpose, appropriate capping groups and group constraints were introduced which mimic the partial charge distribution in the linker when ligated to the dyes and nucleotides (Supplementary Fig 5.13). Protocols for linker building, RESP fitting and updating an existing force field with the new parameters can be found at https://github.com/
fdsteffen/gromacs-tools/fluordynamics/fluorlabel. Dyes and linkers were fused to the RNA at the desired position using a custom-written PyMOL plugin named FluorLabel. The tool supports various labeling chemistries for nucleic acids ranging from 3′/5′-end labeling to amino modified uracils/thymines and ethenoadducts for adenines/cytosines.

5.4.10 Molecular dynamics simulations

Molecular dynamics simulations were run with Gromacs 5.1[325] using AMBER-ff14[326] which includes parmbsc0[327] and χOL3[328,329] corrections for RNA and ϵ/ζOL1[368], χOL4[369] and βOL1[370] corrections for DNA. Ion parameters for K^+ and Cl^- are taken from Joung and Cheatam[330]. The fluorophore labeled RNA was solvated in a TIP3P

dodecahedral water box, charge neutralized with KCl and equilibrated at 300 K and 1 bar with the velocity rescale thermostat and the Parrinello-Rahman barostat respectively. Bonded interactions were constrained by the LINCS algorithm[371] using an integration time step of 2 fs while non-bonded interactions were treated with the Verlet scheme and a cut-off of 1.4 nm. Long-range electrostatics were computed with the particle mesh Ewald method.[372] Distances between the central carbon atoms in the polymethine chain of the donor and acceptor fluorophore were calculated every picosecond. Likewise, the time-dependent orientation factor $\kappa^2(t)$ was computed for each snapshot using `gmx dyecoupl`[343]

$$\kappa^2(t) = (\cos\theta_\mathrm{T} - 3\cos\theta_\mathrm{D}\ \cos\theta_\mathrm{A})^2 \tag{5.19}$$

where θ_D and θ_A are angles between the vector connecting the central carbons and the transition dipole of donor or acceptor respectively, and θ_T is the angle between the two transition dipoles.[166]

5.5 Supplementary Information

Table 5.1 | **DNA sequences used as fluorescent antisense probes for native PAGE.**
The DNA probes are complementary to parts of the expression platform or to the wild type *E. coli* sequence extending the *btuB* riboswitch at the 5′-end (three point mutations, C-6G, U-7A, A-8G, were previously introduced in this segment to prevent interactions with the expression platform).[245]

name	region on RNA	sequence
5-′Cy5-ext	extension platform	(Cy5)-AAG TGC TCG CTA CAG GTT GTA ACA AGT TAC TGT CCA ATA AC-(biotin)
3-′Cy3-exp	expression platform	CGT AAA GCA TCC ACA ATA G-(Cy3)

Table 5.2 | **Single-molecule distance measurements on double-stranded DNA.** Mean FRET efficiencies $\langle E \rangle$ and associated FRET-averaged distances $R_{\langle E \rangle}$ are determined for two different FRET pairs (Atto550/Atto647N and Cy3/Cy5) by confocal spectroscopy. Values for Atto550/Atto647N from the community benchmark study[191] are given as reference. The Förster radius for the dye pairs are $R_0 = 63\,\text{Å}$ (Atto550/Atto647N) and $R_0 = 54\,\text{Å}$ (Cy3/Cy5). The relative ratio $R_{\text{rel}} = \frac{R_{\langle E \rangle,\text{lo}}}{R_{\langle E \rangle,\text{mid}}}$ is calculated as a means to compare different dye pairs independently of R_0.[191]

sample	**position**	$\langle \boldsymbol{E} \rangle$	$\boldsymbol{R}_{\langle \boldsymbol{E} \rangle}$ (Å)	R_{rel}
1-lo[191]	dT31-Atto550 – dT31-Atto647N	0.15 ± 0.02	83.6 ± 2.1	1.39
1-mid[191]	dT23-Atto550 – dT31-Atto647N	0.56 ± 0.03	60.1 ± 1.2	1
1-lo[191]	dT19-Atto550 – dT31-Atto647N	0.76 ± 0.015	51.7 ± 0.7	0.86
1-lo	dT31-Atto550 – dT31-Atto647N	0.15	83.6	1.41
1-mid	dT23-Atto550 – dT31-Atto647N	0.58	59.3	1
1-hi	dT19-Atto550 – dT31-Atto647N	0.76	51.7	0.87
Cy-lo	dT31-Cy3 – dT31-Cy5	0.16	71.2	1.23
Cy-mid	dT23-Cy3 – dT31-Cy5	0.40	57.8	1
Cy-hi	dT19-Cy3 – dT31-Cy5	0.58	51.2	0.89
Cy-lo (ACV) [a]	dT31-Cy3 – dT31-Cy5	0.07	84.1	1.44
Cy-mid (ACV) [a]	dT23-Cy3 – dT31-Cy5	0.39	58.3	1
Cy-hi (ACV) [a]	dT19-Cy3 – dT31-Cy5	0.62	49.7	0.85

[a] ACVs are calculated with dye parameters given in Supplementary Table 5.3. The stacked fraction is estimated from the residual anisotropy ($r_\infty = 0.03$ for dT31-Atto550; $r_\infty = 0.02$ for dT31-Atto647N; Supplementary Fig. 5.7) and set to 8% and 5% respectively. The thickness of the CV is set to be equal to R_2. As expected from the low residual anisotropy, reweighting of the AV has no significant impact on the FRET prediction and ACV and AV calculations yield the same result in this case.

Table 5.3 | **Dye and linker dimensions chosen for accessible volume calculations.**
Individual distance estimates for base, spacer and dye components are given in Fig. 5.12. All distances are rounded to the nearest 0.5 Å.

sample	**position**	**labeling chemistry**	**attach. atom**	$\boldsymbol{L}_{\text{linker}}$ (Å)	$\boldsymbol{W}_{\text{linker}}$ (Å)	$\boldsymbol{R}_{\text{dye,1}}$ (Å)	$\boldsymbol{R}_{\text{dye,2}}$ (Å)	$\boldsymbol{R}_{\text{dye,3}}$ (Å)
DNA ruler [a]	dT19/dT23/dT31-C6-Atto550	deoxythymidine	C7	21	3.5	7.0	3.5	1.5
DNA ruler [a]	dT31-C6-Atto647	deoxythymidine	C7	21	3.5	6.8	4.5	1.5
DNA ruler [a]	dT19/dT23/dT31-C6-Cy3	deoxythymidine	C7	21	3.5	8.0	3.0	1.5
DNA ruler [a]	dT31-C6-Cy5	deoxythymidine	C7	21	3.5	9.5	3.0	1.5
btuB	G240-Cy3	3′-hydrazide	C3′	10.0	3.5	8.0	3.0	1.5
btuB	A35/A213-Cy5	ethenoadenosine	N1 or P [c]	18.5	3.5	9.5	3.0	1.5
RBS hairpin [b]	G240-Cy3	3′-phosphate	O3′	20.5	4.5 [d]	8.0	3.0	3.5 [d]
RBS hairpin [b]	U213-Cy5	uridine	C5 (C7) or P [c]	21	4.5 [d]	9.5	3.0	3.5 [d]

[a] see Supplementary Figs. 5.6 and 5.7
[b] see Supplementary Fig. 5.9
[c] in uridine the distance is measured from the introduced methyl group (C7), like in deoxythymidine; if the target nucleoside is involved in a base-pair, choosing the phosphorous as the attachment atom for the ACV calculation may give better results (mimicking a base that swaps outwards to expose the dye).
[d] a slight increase in W_{linker} and R_3 is needed to prevent the dye from sampling an area that is sterically inaccessible.

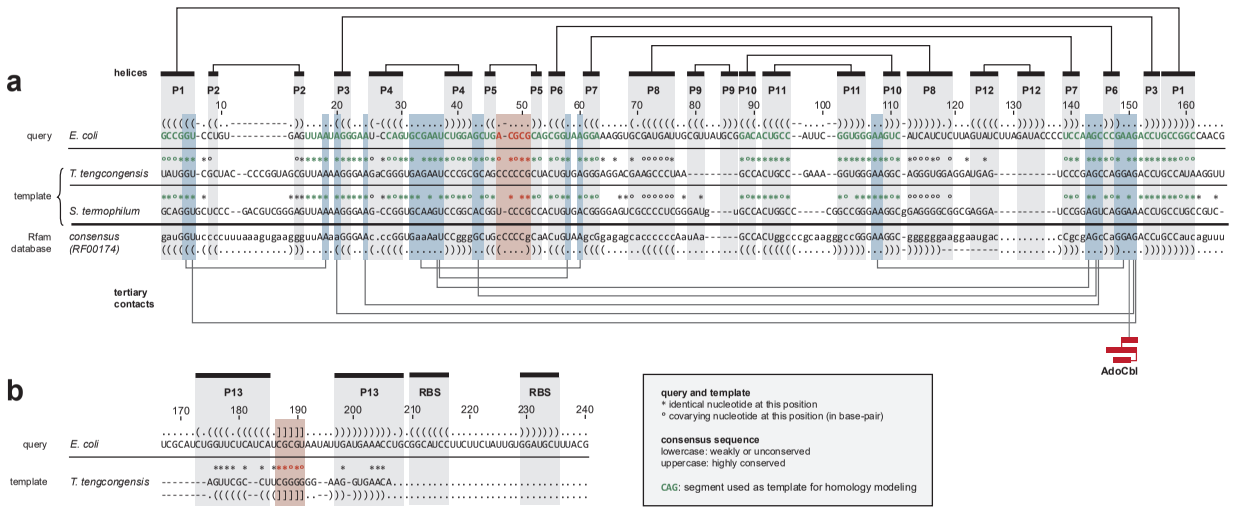


Fig. 5.4 | **Multi-sequence alignment of cobalamin riboswitches.**
a Alignment of the *btuB* riboswitch from *E. coli* to the aptamer consensus sequence (Rfam entry: RF00174). Indicated alongside the *E. coli* riboswitch (query) are the sequences of *T. tengcongensis* and *S. thermophilum*. These two AdoCbl riboswitches have been crystallized and are used as a template for the homology modeling of the *E coli* aptamer (nucleotides highlighted in green). The numbering follows the *E. coli* sequence. Helical regions are shaded in gray (P1-P13) while non-Watson-Crick base pairs are colored blue. The pseudoknot interaction is marked in red. The adenosyl ligand of coenzyme B_{12} contacts A162 in the *T. tengcongensis* structure which corresponds to A150 in the *E. coli* riboswitch. **b** Alignment of the P13 and the pseudoknot in the expression platforms of the *E. coli* and *T. tengcongensis* riboswitches. The ribosomal binding site (RBS) hairpin is not present in the crystal structure and is therefore modeled *de novo*.

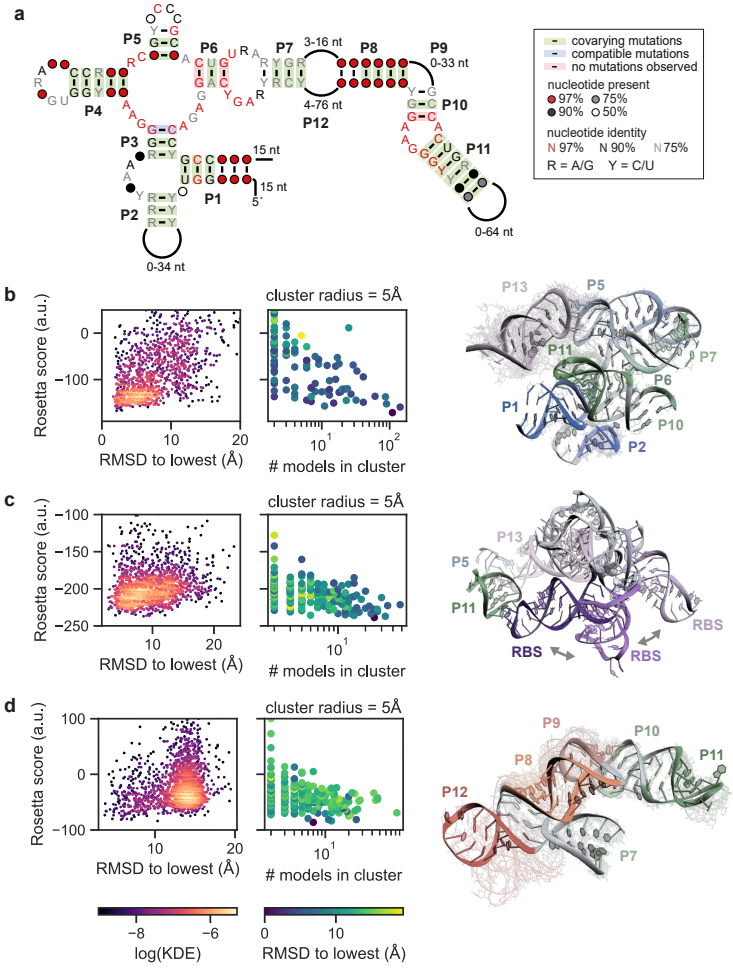


Fig. 5.5 | **Fragment assembly with Rosetta in three phases.**
a Consensus secondary structure from a multi-sequence alignment (Rfam: RF00174) of cobalamin riboswitches annotated with R2R.[373] **b** *De novo* modeling of the core regions P2, P5, P11 and P13. Rosetta energy score ("res4") against the root-mean square deviation (RMSD) of all decoys with respect to the lowest-energy structure. Additionally, the 500 lowest-energy structures were clustered with a radius of 5 Å. The lowest-energy models also tend to have the highest number of neighbors within a cluster, indicating that the models are converging. The ten lowest-energy structures and the chosen model are overlaid. (*continued on next page*)

Fig. 5.5 | (*continued*) **c** *De novo* modeling of the RBS hairpin and its connection to P1. The RBS stem loop is fairly dynamic and adopts multiple conformations among the five lowest energy structures (the three most divergent are depicted). They range from a coaxially stacked hairpin (light violet) via a kinked variant (violet) to a conformation where the RBS docks on the surface exposed P11. **d** Modeling of the peripheral regions P8, P9 and P12. Particularly P12 appears to be dynamic without a Mg^{2+} ions coordinating at the base of the stem.[345] The absence of this divalent ion in the modeling would also help to explain why there is a significant subset of models with a RMSD > 10 Å from the lowest energy structure.

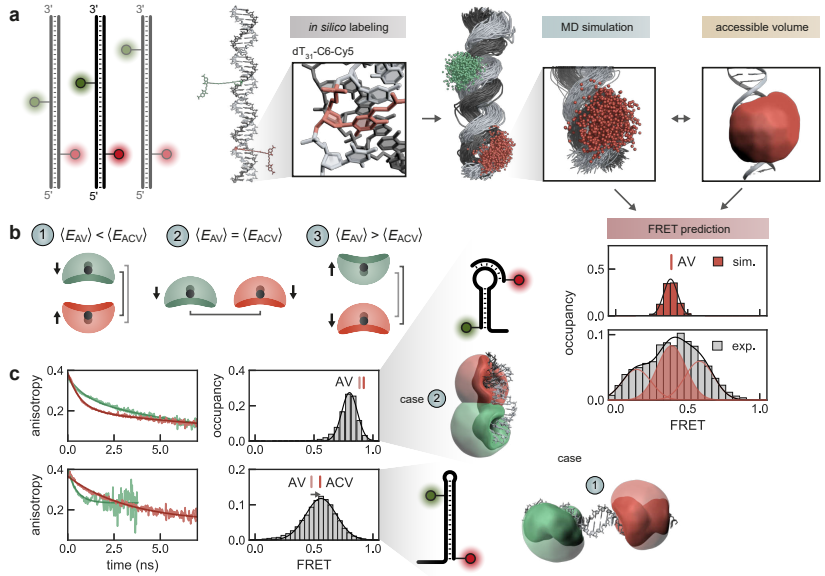


Fig. 5.6 | **Benchmark for FRET based distance measurements.**
a Double-stranded DNA was labeled *in silico* with Cy3 or Cy5 using a modified deoxythymidine and a 6-carbon spacer. An MD simulation generated a structural ensemble. The position of the central C-atom of the carbocyanine fluorophore was saved every 100 ps creating a point cloud of dye positions. The accessible volume approximates the dye distribution for a single static structure. Photon emission events were simulated on top of the MD trajectory (τ_D = 1 ns, τ_A = 1.4 ns, ϕ_D = 0.2, ϕ_A = 0.3) and were binned into a FRET histogram. The mean FRET efficiency $\langle E \rangle$ from the static (AV) and dynamic (MD sim.) was compared to the confocal experiment (dT23Cy3-dT31Cy5, shaded in red). **b** The accessible contact volume (ACV) reweights the grid points by the residual anisotropy which is a measure for interactions of the dye with the biomolecule. In the ACV model the mean dye position (MP) is pulled towards the biomolecular surface, thus affecting also the mean FRET efficiency $\langle E \rangle$. Schematics of the limiting cases are shown. Deviations from the uniform AV are most prominent if the clouds are pointing away (case 1) or facing each other (case 3). **c** Examples demonstrating the influence of the contact volume. Residual anisotropy indicates stacking of the fluorophores with the nucleic acids. In the EBS1*/IBS1* hairpin[178] the donor and acceptor dye clouds are oriented parallel to each other and reweighting has no significant impact on the inter-dye distance. In a second DNA stem loop the ACVs are aligned coaxially and reweighting matches $\langle E_{ACV} \rangle$ with $\langle E_{exp} \rangle$. FRET corrections were applied as described in the Methods section.

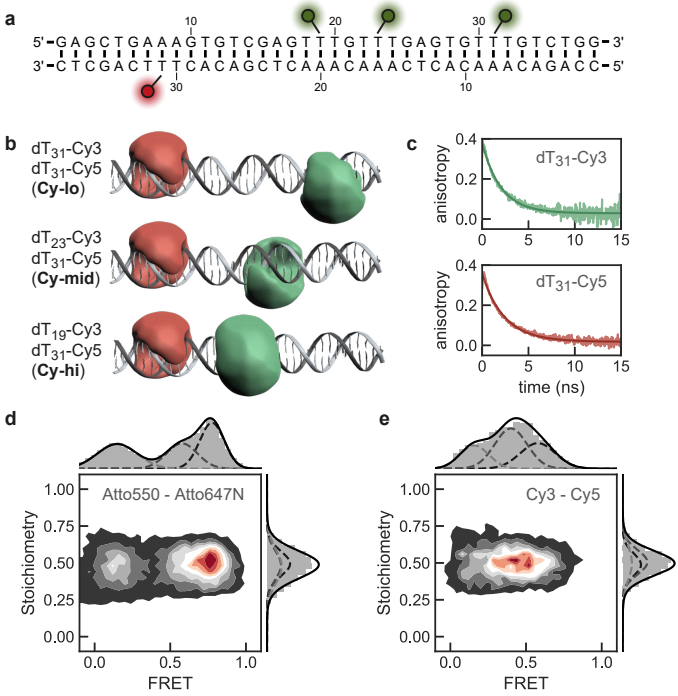


Fig. 5.7 | **Comparison of different FRET labels for distance measurements on double-stranded DNA.**
a Donor and acceptor strand sequences (one donor fluorophore per strand). **b** Accessible volumes calculated for dT31-Cy5 and the three different donor positions dT19-Cy3 (Cy3-lo), dT23-Cy3 (Cy-mid) and dT31-Cy3 (Cy-hi). **c** Anisotropy decays at the labeling position dT31 show nearly no stacking interactions, indicating that contact volume reweighting will not significantly affect the FRET prediction ($\langle E_{\mathrm{ACV}} \rangle = \langle E_{\mathrm{AV}} \rangle$). **d** Merged FRET histograms of the three Atto-labeled DNA helices recorded at 20 mM Mg^{2+} and 5 mM NaCl at pH 7.5 as in the original benchmark.[191] **e** FRET histograms recorded under the same conditions with cyanine labeled DNA. Corrections were applied as described in the Methods section.

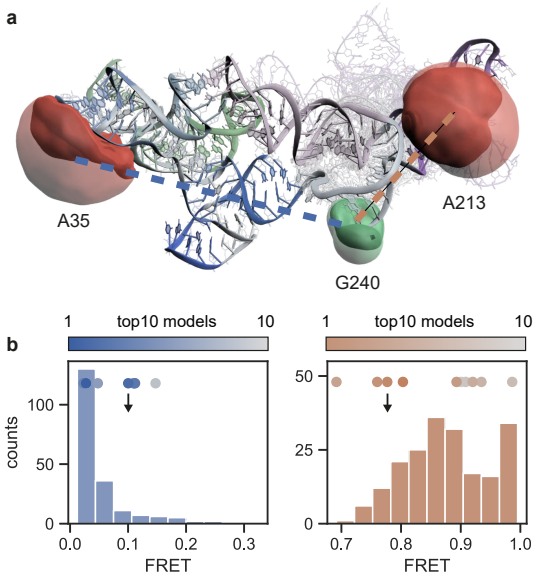


Fig. 5.8 | **Multiple accessible volume screen to filter the lowest-energy Rosetta models.**
a Accessible volumes were calculated for the 200 lowest-energy models at positions A35, A213 and the 3′-end (G240). Weights of the contact volume were determined from the residual anisotropies of the dyes[245] (3′-end: $r_\infty = 0.35$, $w_{CV} = 0.92$; A35: $r_\infty = 0.22$, $w_{CV} = 0.58$; A213 $r_\infty = 0.24$, $w_{CV} = 0.63$) **b** FRET prediction from the ACVs using a Förster radius $R_0 = 54$ Å. The top 10 models are marked by color-graded dots and the depicted model is indicated by an arrow.

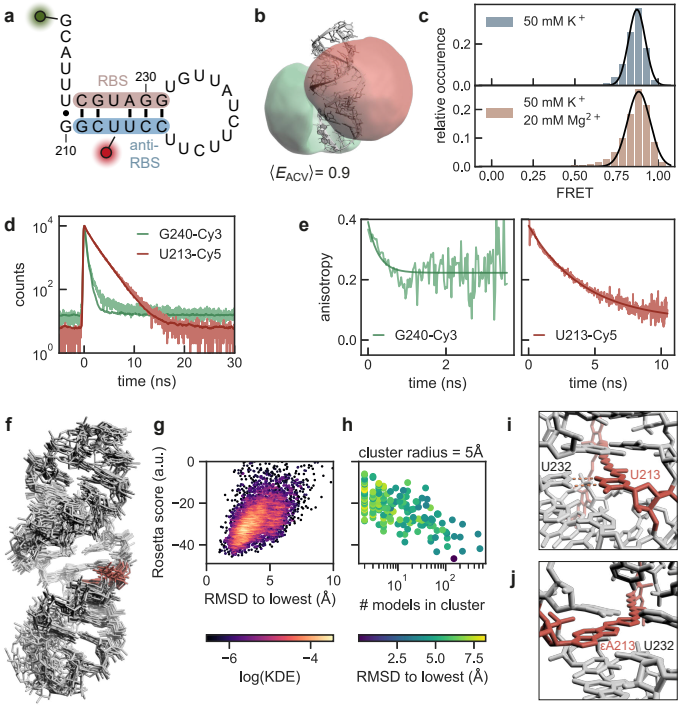


Fig. 5.9 | **The isolated RBS hairpin forms independently of Mg^{2+}.** **a** Secondary structure of the RBS hairpin labeled at the 3′-end with Cy3 and on U213 with Cy5. **b** Accessible-contact volumes (ACV) predict an average FRET value of 0.9 (the contact volume is omitted for clarity). **c** FRET histograms in the presence and absence of Mg^{2+}. Possible traces of divalent ions (histograms recorded without EDTA) were estimated to be $< 1‰$, thus micromolar or lower. **e** Ensemble fluorescence lifetime decays of G240-Cy3 and U213-Cy5. The lifetime of Cy3 is shortened by FRET. **f** Fluorescence anisotropy decays indicate stacking interactions of Cy3 with the single-stranded overhang ($r_{\infty,\mathrm{Cy3}} = 0.22$; $r_{\infty,\mathrm{Cy5}} = 0.06$). Weights of the CV were calculated using Equation 5.10 ($w_{\mathrm{CV,Cy3}} = 0.57$; $w_{\mathrm{CV,Cy5}} = 0.17$). **f** Ten lowest-energy structures predicted with FARFAR2 (U213 is highlighted in red). **g** Rosetta score ("res4") versus RMSD to the lowest-energy model. **h** Rosetta score versus the number of models per cluster (after clustering the 500 lowest-energy decoys with a radius of 5 Å). The number of neighbors within a cluster was previously identified as a model selection metric.[156,164] **i** Cy5 dye attached to the Hoogsteen edge of an amino-modified uracil leaving the WC face free to interact.[120] **j** The ethenoadduct on adenine blocks the WC edge and the nucleotide needs to turn around to expose its Hoogsteen edge for possible interactions (schematic illustration).

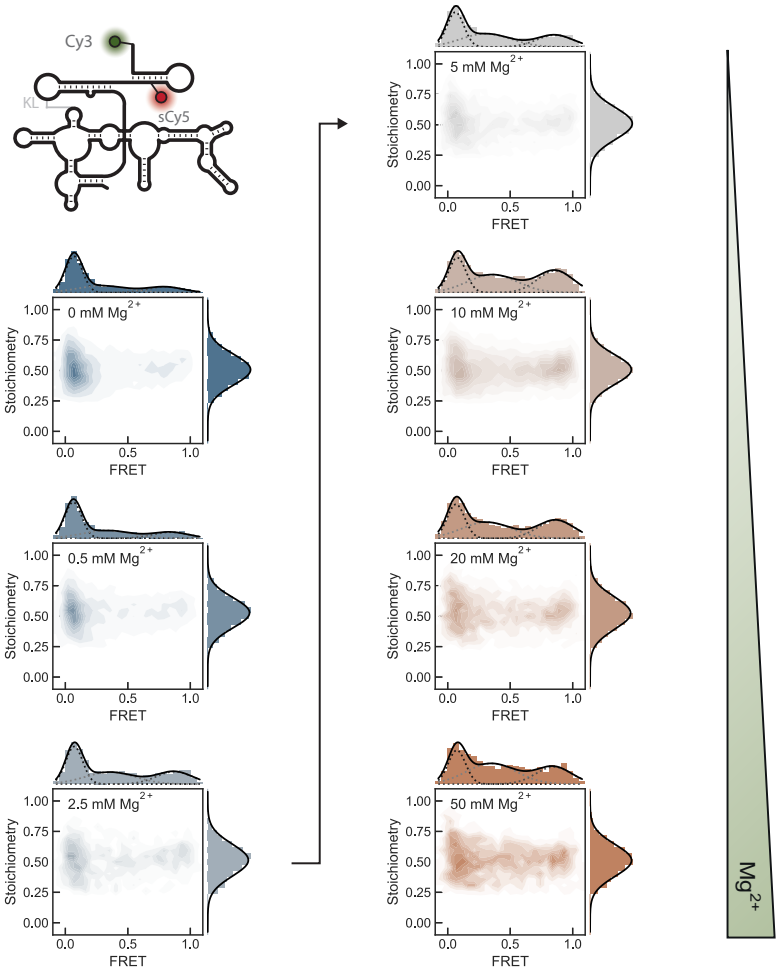


Fig. 5.10 | **Mg^{2+} titration along the 3'-A213 reaction coordinate indicates formation of the RBS hairpin.**
2D stoichiometry-FRET histograms with donor-only and acceptor contributions removed by a minimum red photon threshold $F^{\mathrm{A_{em}}}_{\mathrm{A_{ex}}} > 50$ photons and a stoichiometry offset of 0.2. Blinking was removed with a ALEX-2CDE< 20.[189] Corrections were applied as described in the Methods section.

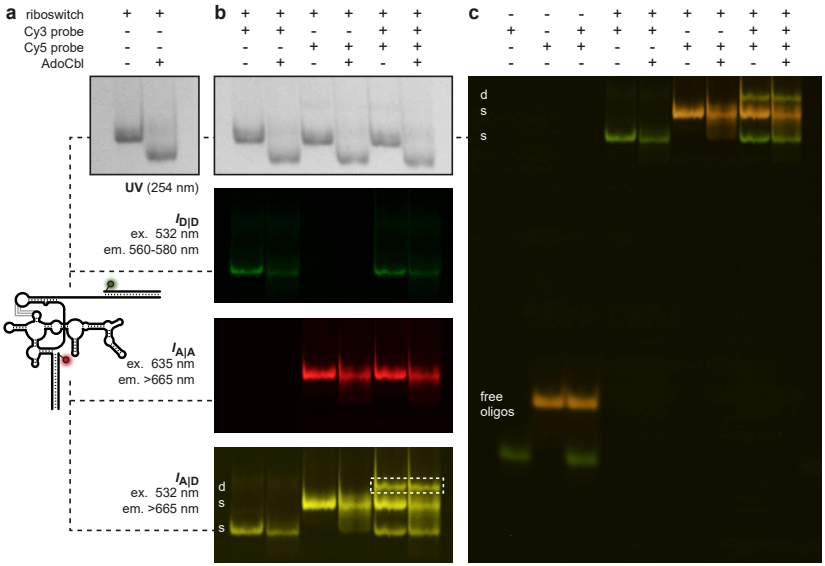


Fig. 5.11 | **The btuB switches conformation in response to AdoCbl.**
a Native polyacrylamide gel electrophoresis (PAGE) shows a band shift when the *btuB* riboswitch is incubated with AdoCbl in the absence of Cy3/Cy5 DNA probes. **b** In the presence of the labeled DNA oligonucleotides, the RNA still switches conformation as seen under UV illumination but only the upper band is fluorescently labeled. One reason could be that the Cy3 DNA antisense oligonucleotide anneals to the expression platform, blocks the RBS hairpin and prevents switching of any labeled RNA. Alternatively, compaction of the RNA may bring Cy3 in close proximity to AdoCbl which quenches the fluorescence.[74] **c** Full view of the gel with all channels overlaid. The RNA was added in tenfold excess with respect to the DNA oligos to ensure complete saturation. The gel was recorded in the presence of 3 mM $MgCl_2$ and 50 mM KCl at pH 7.5.

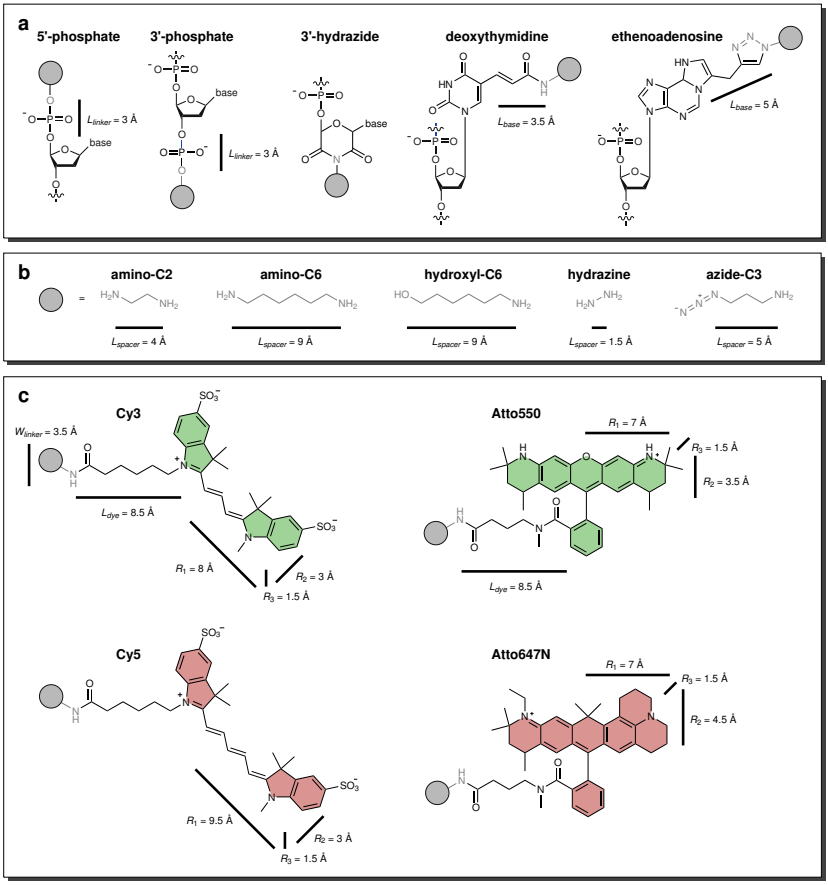


Fig. 5.12 | **Fluorophore parameters for ACV calculations.**
a Coupling chemistry at the nucleobases. Distances are measured in PyMOL and rounded to the nearest 0.5 Å. **b** Spacers increase the dye flexibility and are decorated by different functional groups. **c** Donor (green) and acceptor (red) fluorophores parameterized by three radii R_1, R_2 and R_3. The maximally extended linker length L_{linker} is calculated as the sum of L_{base}, L_{spacer} and L_{dye}. In certain situations, it may be necessary to increase W_{linker} and R_3 slightly to prevent the dye from exploring unrealistic areas (i.e. narrow cavities).

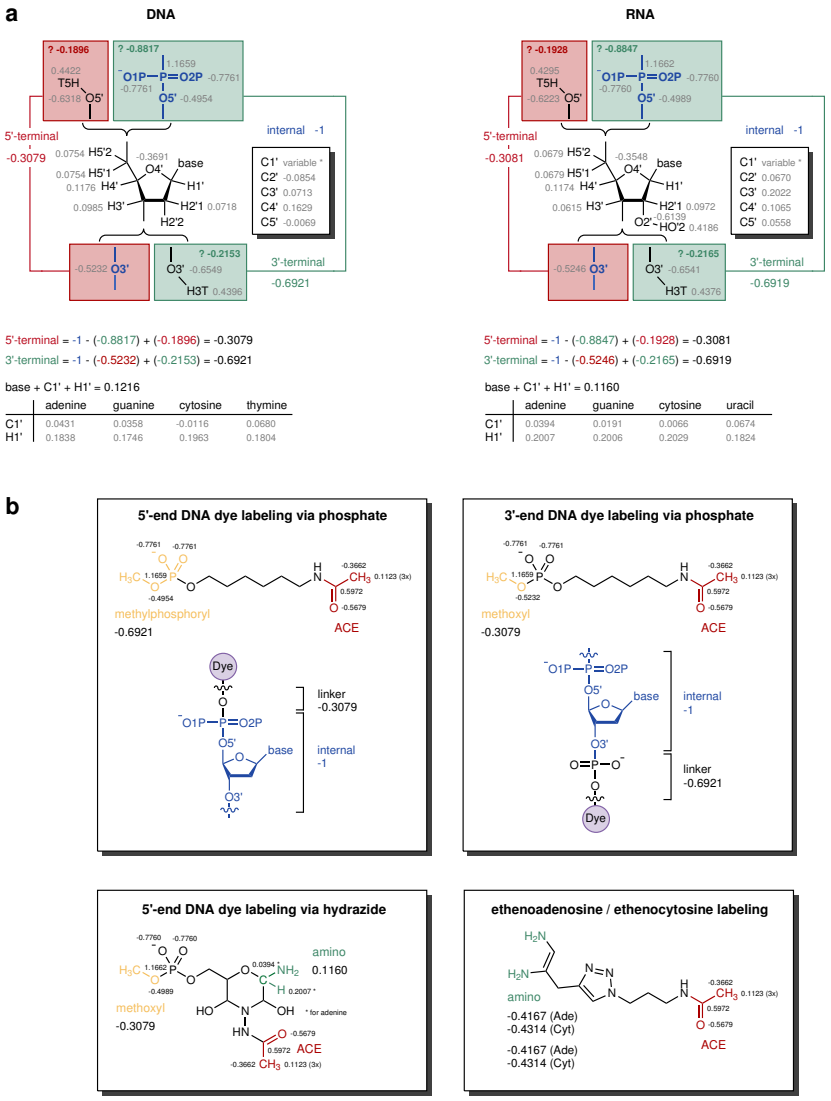


Fig. 5.13 | **Force field parameterization for dye-linker atoms.**
a Partial charges of RNA and DNA residues in the AMBER force field according to Cornell et al.[326] **b** Partial charges for the capping groups of the dye linker constrained during RESP fitting.[366]

Chapter 6

Summary and Outlook

The energy landscape of nucleic acids has often been described as a "rugged" mountain scenery. RNA folding is an attempt to cross those lands. There are many parallel routes through the energy valleys and over the ridges. Metal ions, proteins and metabolites interact with the RNA along the way, guiding it towards its native state or into kinetic traps.

This analogy illustrates how heterogeneous and non-linear the folding paths of RNA are. This thesis explored such energy landscapes and tried to reconstruct their topology using a variety of biophysical techniques and most importantly fluorescence spectroscopy. Fluorophores are our reporters and positioning them at the right spots is key to an informative reaction coordinate. In chapter 3 we reviewed the most popular labeling strategies of RNA for single-molecule applications.[314] Currently available methods are either limited by the size of the target molecule (solid-phase chemical synthesis with modified phosphoramidites[233]), restrictive regarding possible dye locations (3'/5'-end labeling with T4 ligase or polynucleotide kinase) or induce major changes to the RNA (DNA probe hybridization[247]). Intending to mitigate some of these shortcomings, we designed a post-transcriptional labeling approach for RNAs of arbitrary size. It uses nucleic acid templated chemistry in order to site-specifically transfer a dye-reactive functional group (here an alkyne) onto a pre-selected adenine or cytosine residue (Fig. 6.3). An azide dye then reacts with the alkyne in a Click reaction. If desired, a second dye can be coupled to the 3′-terminus concurrently using an orthogonal reaction. In the future, the functional groups could be inverted (azide-RNS and alkyne-dye) or expanded to other coupling chemistries (e.g. maleimides) and different sites on the nucleotide (2′-OH, A/G-N7, C-C5). While the former would make it possible to attach dyes at multiple internal positions concomitantly, the latter would allow to leave the Watson-Crick edge untouched and thus further improve labeling of duplex regions.

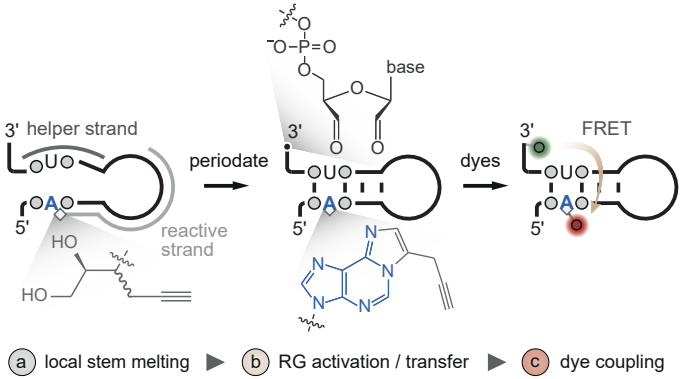


Fig. 6.1 | **Schematic of our site-specific, fluorescent labeling strategy using DNA-guided chemistry.**
We introduced a new site-specific labeling technique developed for long RNAs such as ribozymes and riboswitches, which are inherently difficult to label internally because of their size.

The combination of FRET with single-molecule detection was the method's ticket to biology. With the possibility of a subpopulation-specific analysis, single-molecule experiments have the edge over classical ensemble assays. The high dynamic range of fluorescence from picoseconds to minutes has allowed to measure kinetics on very different timescales.[255] It has unveiled short-lived intermediates as well as heterogeneities among conformers and their interconversion rates.[175,177,282] A complication for analyzing kinetics arises if there is insufficient spatial separation among FRET coordinates. Such overlapping FRET states are hence referred to as degenerate. With some informed biochemical guess about probable or also structurally incompatible connectivities, it is still possible to infer (parts of) the state network by hidden Markov modeling on a set of fluorescence time traces. Chapter 4 explores an RNA tertiary contact found at the core of a group II intron which displays heterogeneous unbinding rates due to state degeneracy in the bound conformation (Fig. 6.2).[178] NMR chemical shifts assays have previously identified several coordination sites for Mg^{2+} ions.[280] Their transient, and thus incomplete, occupation was found as a main source of heterogeneity in exon dissociation rates. Simultaneous binding of the exon and a metal ion induce longer dwell times on the RNA hairpin while exon-intron contacts lacking reinforcement by Mg^{2+} tend to dissociate faster.

The ion thus introduces a bottleneck that stalls the flow in the unbinding pathway. Conversely, heterogeneity is largely abrogated in RNA/DNA hybrid contacts and dissociation is faster by about one order of magnitude. With a combination of total correlation spectroscopy (TOCSY) NMR experiments and molecular dynamics we found that RNA/DNA hybrids show fast sugar puckering which is thought to weaken the binding of the two strands.

The exon and intron binding site interaction (EBS1/IBS1) studied here in isolation serves as a proxy for exon recognition by group II introns. Undoubtedly, this model system lacks some of the architectural details that make up the ribozyme's active site, yet some important conclusions on the timing of exon binding and catalysis can still be drawn and shall be outlined in the following. To maximize turnover an efficient enzyme binds its substrate with sufficient affinity but readily releases the product once it has formed. Likewise, ion binding and the sugar puckers are suggested to fine-tune the residence time of the exon in the core of the ribozyme just long enough for the two transesterification reactions to take place. Both a binding that is too weak and one that is too tight would prevent efficient catalysis and result in splicing deficits. In the first case, the RNA does not recognize its target sufficiently and the intron is retained. When the 5′-exon dissociates after the first splicing step, the exons remain unligated. In the second case, where binding is too strong, discrimination between mismatched and cognate sites is impaired (no proofreading), which is particularly detrimental to genome integrity when introns are mobile (retrotransposition).[278,374] To test some of these hypotheses, future work could build upon the herein established reaction coordinate (EBS1/IBS1 binding) and internal labeling workflows (ethenoadenosines/cytidines) to monitor exon binding in *cis* or *trans* in the context of the whole intron and relate it to the timescales of catalysis.

An important puzzle piece to a quantitative interpretation of FRET experiments are suitable dye models. They estimate the uncertainty introduced by the flexible, fluorescent reporters and thus link a measured FRET signal with a biomolecular structure. Carbocyanine fluorophores, the most popular dyes for labeling RNA, are prone to interact with exposed nucleotides.[167] Stacking contacts increase the brightness of dyes by a preferential stabilization of the fluorescent *trans* state. This phenomenon is called RNA-induced fluorescence enhancement (RIFE)[171] or PIFE[170,375] in case of proteins. The tendency of these fluorophores

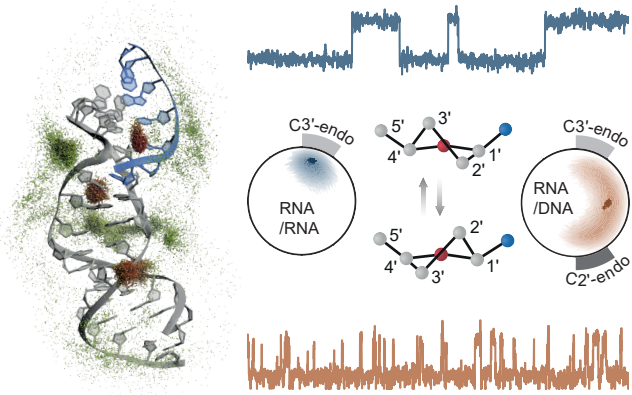


Fig. 6.2 | **A sugar switch links kinetic heterogeneity to structure.** We combined single-molecule FRET and molecular dynamics simulations to elucidate the heterogeneous unbinding kinetics of RNA and DNA tertiary contacts in which metal binding and sugar puckering are constantly reshaping the energy landscape.

to stick to the surface of RNA led us to partition the sterically accessible volume (AV) of an established dye-linker model[258,362] into two parts, a contact and a free volume. Both point clouds are assigned a weight based on the residual fluorescence anisotropy which shifts the mean position of the fluorophore within the accessible volume and refines the predicted FRET efficiency. In chapter 5 we integrated this dye model into a homology and *de novo* RNA structure prediction workflow. As a target we selected a coenzyme B_{12} riboswitch of *E. coli*. As one of the largest riboswitches discovered to date, the *btuB* leader sequence has not been amenable to crystallization. However, the structures of two homologs were previously solved and served as blueprints for modeling the *E. coli* riboswitch. To probe the accessibility of the ribosomal binding site, we labeled this motif with two fluorophores. Addition of Mg^{2+} prepares the riboswitch for interaction with the metabolite in two ways: it induces a global compaction of the RNA through formation of a pseudoknot and locally increases the fraction of folded RBS hairpins. We speculate that such coordinated folding is crucial to reach a binding competent structure. Future work will extend the existing FRET network to include labels near the two kissing loops or within P11 in order to test this functional hypothesis.

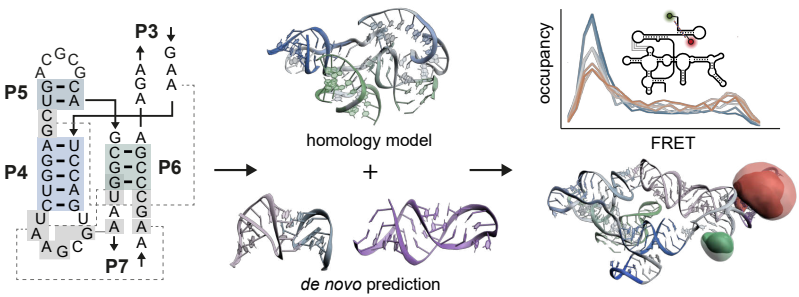


Fig. 6.3 | **Workflow of FRET-assisted homology and de novo prediction of RNA.**
We modeled the aptamer and expression platform of a cobalamin riboswitch and probed the accessibility of the ribosomal binding site hairpin by single-molecule FRET.

More than sixty years have passed since the Francis Crick first formulated what he called the "central dogma" of molecular biology.[17] In its original phrasing[1] the concept is still widely recognized as foundational for the encoding and transmission of information among different alphabets. Within this framework RNA assumes the role of a messenger.[376] Today we know that RNA integrates as a regulatory unit in numerous processes in the cell. The functional diversity of RNA will continue to grow as new and rarer classes of riboswitches and ribozymes are discovered. Herein, two representatives have been studied in greater detail by spectroscopic and computational means. This work contributes to the continuous integration of experiments and simulations. With increasing computational power and temporal resolution the time scales of experiment and simulation will steadily converge and allow us to fill the picture of a fissured energy landscape with quantitative descriptions of dynamic conformational ensembles.

[1] Crick stated it as follows: *"once information has passed into protein it cannot get out."*[17] According to Crick this formulation does not explicitly exclude the passage from RNA to DNA (reverse transcription) as he pointed out in another well-known article in 1970.[18]

Chapter 7

Zusammenfassung und Ausblick

Um die Energielandschaft von Ribonukleinsäuren zu verstehen, lohnt sich ein Ausflug in die Berge. RNS Faltung gleicht einer Wanderung durch Energietäler und über Höhenzüge. Zahlreiche Routen führen durch das Gelände. Mit von der Partie sind Metallionen, Proteine und verschiedene Metaboliten, welche mit der RNS interagieren und so die RNS zu ihrem gefalteten Zustand leiten, bisweilen aber auch in kinetische Fallen tappen lassen.
Diese Analogie soll aufzeigen, wie wichtig das Verständnis um die Topologie ist, um den Faltungsprozess und den Funktionsmechanismus von Biomolekülen zu verstehen. In dieser Arbeit wurden unterschiedliche biophysikalische Messmethoden eingesetzt, um diese Landschaft zu erkunden. Dazu zählt insbesondere die Fluoreszenzspektroskopie. Eine zentrale Aufgabe kommt dabei den Farbstoffen zu, deren Positionen im Biomolekül und zueinander die Reaktionskoordinate bestimmen. In Kapitel 3 wurden die gängigsten Verfahren zur Fluoreszenzmarkierung von Nukleinsäuren für die Einzelmolekülspektroskopie vorgestellt. Die Methoden unterscheiden sich in Bezug auf die Länge und Stelle der RNS, die markiert werden kann und den Einfluss auf die Struktur. Die Festphasen-Synthese, bei welcher ein modifiziertes Nukleosid an jeder beliebigen Stelle eingebaut werden kann, ist auf 50-100 Basen beschränkt. Dieser gegenüber stehen enzymatische Reaktionen mit T4 Ligase oder Polynukleotidkinase, welche unabhängig von der Grösse der RNS sind, jedoch nur die beiden Enden markieren können. Die dritte Möglichkeit stellen Hybridisierungsmethoden dar, bei denen fluoreszent markierte DNS Stränge nicht-kovalent an die RNS gebunden werden. Dies bedingt oftmals eine vorgängige Erweiterung von internen, einzelsträngigen Schleifen, was mitunter Auswirkungen auf Struktur und Funktion des Biomoleküls nach sich ziehen kann.

Mit dem Ziel, einige dieser Unzulänglichkeiten zu beseitigen, haben wir eine posttranskriptionelle und damit längenunabhängige Methode entwickelt, in der zwei Reaktionspartner mittels eines Oligonukleotids nahe zueinander gebracht werden (*nucleic acid templated chemistry*). Zum einen ist dies ein Alkin mit einer zusätzlichen Diolgruppe und zum anderen ein vorgängig ausgewähltes Adenin oder Cytosin auf der RNS (Fig. 6.3). Das Amin der Base reagiert unter Wasserabspaltung mit der zum Aldehyd oxidierten Diolgruppe, wobei das Alkin auf die RNS übertragen wird. Da während der Diolaktivierung der Zucker am 3′-Ende ebenfalls oxidiert, kann hier gleichzeitig ein zweiter Farbstoff über eine orthogonale Reaktion eingesetzt werden. Um mehrere Basen innerhalb der RNS zu markieren, könnten einerseits die Reaktionspartner umgedreht werden. Folglich würde das Azid auf der RNS platziert und der Farbstoff mit dem Alkin gekoppelt. Alternativ könnte die reaktive Gruppe durch eine andere, beispielsweise ein Maleimid, ersetzt werden. Weitere, jedoch deutlich komplexere Optimierungsmöglichkeiten ergäben sich bei der Modifikationsstelle am Nukleosid. Das Ethenoaddukt blockiert die Watson-Crick Basenpaarung, wodurch Duplexregionen zwar markiert werden können, dies aber mit einem Stabilitätsverlust einhergeht. Modifikationen am 2′-OH, dem N7 von Adenin oder Guanin oder dem C5 von Cytosin wären hierfür potenzielle Ansatzpunkte.

Die Einzelmoleküldetektion hat FRET als Methode für die Erforschung von Biomolekülen interessant gemacht. Mit der Möglichkeit, einzelne Konformationen voneinander getrennt zu untersuchen, ist die Einzelmolekülspektroskopie gegenüber klassischen Ensembleexperimenten klar im Vorteil. Der aussergewöhnlich hohe Dynamikumfang der Fluoreszenz, der von Picosekunden bis hin zu mehreren Minuten reicht, erlaubt es, Wechselwirkungen und Bewegungen innerhalb von Biomolekülen auf unterschiedlichsten Zeitskalen zu erfassen.[255] Dabei wurden kurzlebige Konformationen entdeckt und Inhomogenitäten in den Reaktionsraten aufgeklärt.[175,177,282] Die Kinetikanalyse wird allerdings erschwert, wenn sich FRET Zustände überlagern. Weist ein solcher degenerierter FRET Zustand mehrere Relaxationsraten auf, wird dies auch als kinetische Heterogenität bezeichnet. Mit gewissen Vorkenntnissen über wahrscheinliche oder auch strukturell inkompatible Übergänge des zu untersuchenden Systems ist es trotz allem möglich, zumindest Teile des Zustandsnetzwerks über ein verborgenes Markov-

modell zu entschlüsseln. In Kapitel 4 haben wir uns einen Tertiärkontakt angeschaut, welcher Teil des aktiven Zentrums eines Gruppe II Introns ist und mit dem benachbarten 5′-Exon interagiert (Fig. 6.2). Die Dissoziation von Exon und Intron ist heterogen, was auf den Besetzungsgrad von Mg^{2+} Bindungsstellen zurückzuführen ist. Wenn Metallion und Exon zur gleichen Zeit an der RNA Haarnadelschleife gebunden sind, verlängert sich die Aufenthaltswahrscheinlichkeit des Exons am Intron, während ein Kontakt, der nicht durch Mg^{2+} stabilisiert wird, schneller dissoziiert. Wird die Exon-RNA durch eine Exon-DNA ersetzt, erhöhen sich die Dissoziationsraten um etwa eine Grössenordnung und werden gleichzeitig homogener. Anhand von TOCSY NMR Experimenten und Moleküldynamiksimulationen konnten wir zeigen, dass die Desoxyribose des DNA Strangs schnell zwischen mehreren Konformationen hin- und herwechselt (*puckering*), was wahrscheinlich zu einer schwächeren Bindung führt und so die Dissoziation gegenüber RNA/RNA Kontakten begünstigt.
Die hier isoliert untersuchte Exon-Intron Bindungsstelle (EBS1/IBS1) erlaubte eine detaillierte chemische Betrachtungsweise, welche in der Form im Kontext des gesamten Introns schwierig durchzuführen gewesen wäre. Zweifelsohne fehlen dem Modellsystem die Feinheiten des aktiven Zentrums, dennoch ergeben sich aus der Bindungskinetik einige interessante, mechanistische Überlegungen zu Zeitpunkt und Steuerung der Katalyse in Ribozymen. Ein effizientes Enzym, unabhängig ob Protein oder RNS, maximiert seinen Durchsatz, indem es das Substrat mit ausreichender Affinität bindet, das Produkt indes auch leicht wieder freisetzt. Wir vermuten, dass Gruppe II Introns über Magnesiumbindung und Zuckerkonformation die Aufenthaltsdauer des Exons in der aktiven Stelle für die beiden Transesterifikationsreaktionen optimiert haben. Eine zu schwache Bindung würde dazu führen, dass das Intron nicht oder unvollständig herausgeschnitten und die Exons nicht ligiert würden. Im Gegensatz dazu erschwert eine zu starke Bindung die Fehlerkorrektur, wenn ein falsches Exon erkannt wird.[278] Dieses Spleissdefizit hat weitreichende Auswirkungen auf die Genomintegrität, wenn mobile Introns involviert sind, die sich mittels Retrotransposition an anderen Stellen wieder ins Erbgut integrieren können.[374] Um einige dieser Hypothesen zu testen, könnte die EBS1/IBS1 Reaktionskoordinate nun mit Hilfe der in Kapitel 3 vorgestellten Markierungsmethode mit dem ganzen Intron untersucht und diese dann mit der Katalyse in Verbindung gesetzt werden.

Farbstoffmodelle sind für eine quantitative Auswertung von FRET Experimenten von zentraler Bedeutung. Sie verknüpfen das gemessene FRET Signal mit der Struktur, indem sie die Beweglichkeit des Farbstoffs in Bezug auf das Biomolekül abschätzen. Carbocyanine gehören zu den am weitesten verbreiteten Farbstoffen für RNS. Aufgrund ihrer Molekülstruktur haben sie die Tendenz, mit exponierten Basen insbesondere an den Enden und in einzelsträngigen Regionen zu interagieren. Dies führt zu einer höheren Quantenausbeute, was wir in Analogie zum gleichen Phänomen in Proteinen[170,172] als RNS-induzierte Fluoreszenzsteigerung (*RNA-induced fluorescence enhancement*, RIFE[171]) bezeichnen. Diese Beobachtung hat uns dazu veranlasst, den sterisch zugänglichen Raum des Farbstoffs in ein freies Volumen und ein Kontaktvolumen zu unterteilen. Die beiden Punktwolken werden mit der Fluoreszenzanisotropie gewichtet, wodurch sich die gemittelte Position des Farbstoffs leicht verschiebt und die mittlere FRET Effizienz besser bestimmt ist. Im letzten Kapitel haben wir dieses Farbstoffmodell zur Strukturvorhersage eines Riboschalters benutzt. Diese RNS steuert die Expression des *btuB* Gens in *E. coli*, das ein Transportprotein für Coenzym B_{12} codiert. Der *btuB* Riboschalter gehört zu den grössten seiner Art und konnte aufgrund seiner Flexibilität bislang nicht kristallisiert werden. Es existieren jedoch homologe Teilstrukturen von zwei anderen Bakterien, *S. typhimurium* und *T. tengcongensis*, welche als Vorlage benutzt werden konnten. Ein wichtiges Strukturmerkmal für die Genregulation, das in den bisherigen Strukturen fehlte, ist die ribosomalen Bindungsstelle (RBS). Dessen Zugänglichkeit bestimmt, ob das anschliessende Gen vom Ribosom abgelesen werden kann oder nicht. Das Faltungsgleichgewicht dieses Elements haben wir mittels FRET untersucht und dabei herausgefunden, dass Mg^{2+} den Riboschalter in zweierlei Hinsicht beeinflusst. Der Übergang zu einer kompakten Struktur des Aptamers wird begleitet von einer Faltung der Expressionsplattform und damit der RBS Haarnadelschleife. Wir denken, dass ein koordinierter Mechanismus wichtig für die Bindung des Metaboliten ist. Der Einbau von Farbstoffen an weiteren Stellen, so beispielsweise in der Nähe des Pseudoknotens oder in P11, könnte dazu dienen, diese Vermutung experimentell zu belegen.

Etwas mehr als sechzig Jahre sind vergangen, seit Francis Crick seine These zur Übertragung von genetischer Information aufstellte, die er als "Zentrales Dogma" der Molekularbiologie[17] bezeichnete. In seiner ursprünglichen Formulierung[1] hat dieses Konzept weitgehend bis heute Bestand, obschon es sich hinsichtlich der RNS auf dessen Rolle als Vermittler zwischen DNS und Proteinen beschränkt. Heute wissen wir, dass nicht-kodierende RNS an einem breiten Spektrum von zellulären Prozessen beteiligt ist. Zwei solcher RNS Moleküle wurden hier mit spektroskopischen und computergestützten Methoden genauer untersucht. Diese Arbeit leistet einen Beitrag zur kontinuierlichen Integration von Experiment und Simulation. Mit steigender Rechenleistung und Auflösung werden die Zeitskalen in Zukunft noch weiter zusammenrücken. Auf diese Weise wird unser Bild einer hügeligen Energielandschaft zunehmend mit quantitative Beschreibungen dynamischer Konformationsensembles gefüllt.

[1] In den Worten von Crick: *"once information has passed into protein it cannot get out."*[17] Diese Formulierung schliesse die Informationsübertragung von RNA zu DNA (reverse Transkription) nicht grundsätzlich aus, wie er später betonte.[18]

Acknowledgment

First and foremost, I would like to thank my PhD advisor Prof. Roland Sigel for giving me the freedom to pursue my ideas in his lab first as Master and then as PhD student. I am grateful for your constant support over the years including your active encouragement to go and present our work on conferences in Europe and overseas.

I would also like to thank Prof. Ben Schuler (University of Zurich) and Prof. Helmut Grubmüller (Max Planck Institute for Biophysical Chemistry, Göttingen) for their valuable advice as members of my PhD committee.

Special thanks are directed to Prof. Richard Börner for being an excellent mentor. Thank you for introducing me to the field of single-molecule spectroscopy and for many hours of scientific exchange. It has been an exciting journey through the realms of physics and biology.

Furthermore, I would like to express my thanks to

Dr. Sofia Gallo and Dr. Michelle Schaffer for introducing me to the world of RNA and riboswitches;

Dr. Meng Zhao for sharing thoughts and expertise on the fluorescent labeling of nucleic acids;

Dr. Silke Johannsen for many valuable inputs on RNA structure and NMR spectroscopy;

Dr. Susann Zelger-Paulus for stimulating conversations on the biochemistry of group II introns;

Dr. Mauro Schilling for countless discussions over coffee on chemistry and molecular simulations;

Dr. Richard Cunha for teaching me the tricks and tweaks of Plumed;

Dr. Mélodie Hadzic for puzzling together over MASH-FRET;

Dr. Mokrane Khier and Dr. Danny Kowerko for a fruitful collaboration on EBS/IBS.

I would like to extend my thanks to the present and former members of the Sigel-Freisinger laboratory whom I had the pleasure to work with, including Dr. Fabio Amadei, Kenneth Adea, Besim Fazliji, Esra Ahunbay, Maya Gulotti and Prof. Eva Freisinger.

My lab mates Zenghui Wang, Ilaria Przytula-Mally and Moritz Durtschi made experiments in F-78 a pleasant experience.

I would also like to thank Ramona Erni for her untiring commitment to keep things running smoothly in the Department.

Financial support from the UZH Forschungskredit (FK-17-098) and the CMSZH graduate school is gratefully acknowledged. The continuing efforts of the graduate school in organizing the instructive transferable skill courses is also highly appreciated.

Final thanks go to my parents and to my brother for their support during my PhD.

Fabio Steffen
July, 2020

Curriculum vitae – Fabio Steffen

Personal

work address:	University of Zurich, Department of Chemistry Winterthurerstrasse 190, 8057 Zurich
e-mail:	fabio.steffen@chem.uzh.ch
web:	www.chem.uzh.ch/en/sigel/members/steffen
ORCID:	0000-0001-8795-2212

Education and Work experience

PhD in Biophysics 01.2016 – 09.2020	**University of Zurich** | Prof. Dr. *Roland Sigel* Single-molecule FRET guided modeling of RNA structure and dynamics
M.Sc. Biochemistry 09.2014 – 08.2015	**University of Zurich** | Prof. Dr. *Roland Sigel* A comparative analysis of carbocyanine photophysics in the framework of nucleic acids
Research Intern 02.2014 – 03.2014	**University Hospital ZH** | Prof. Dr. med. *Reto Schüpbach* Signal transduction by human protease-activated receptors
B.Sc. Biochemistry 09.2011 – 07.2014	**University of Zurich** | Prof. Dr. *Amedeo Caflisch* Small-molecule based inhibition of bromodomains
Matura 09.2011 – 07.2014	**Kantonsschulen Winterthur / Nyon** | Latin / English Un joyau linguistique menacé de disparition – Le patois francoprovençal en quête de revitalisation

Teaching and Outreach

2016 – 2017	**Laboratory course in Chemistry** | Undergraduates in Medicine
2015 – 2019	**Children University** | Chemistry experiments for school children
2013 – 2014	**Excercise sessions in Chemistry** | Undergraduates in Biology

Conference Talks

2018	**Swiss Chemical Society** (Lausanne, CH)
	Fluorescent Biomolecules and Building Blocks (Glasgow, UK)
	RNA society meeting (Berkeley, USA)
2017	**Methods and Applications in Fluorescence** (Bruges, BE)
	Swiss Chemical Society (Bern, CH)

Awards and Fellowships

2019	**Poster Award**, SCS Fall meeting
2018	**Best Oral Presentation**, SCS Fall meeting
2017	**Forschungskredit**, University of Zurich
2017/18/19	**Travel Awards**, Merck, Bruker and RNA society
2015	**Semester prize**, Faculty of Science, UZH

Publications

2020 **Nature Communications** 2020, *11*, 104
F. D. Steffen*, M. Khier*, D. Kowerko, R. A. Cunha, R. Börner, R. K. O. Sigel *shared first authorship
doi.org/10.1038/s41467-019-13683-4

2019 **Chimia** 2019, *73*, 257-261
F. D. Steffen, R. Börner, E. Freisinger, R. K. O. Sigel
doi.org/10.2533/chimia.2019.257

2018 **Nucleic Acids Research** 2018, *46*, e13
M. Zhao, F. D. Steffen, R. Börner, M. F. Schaffer, R. K. O. Sigel
doi.org/10.1093/nar/gkx1100

2016 **Physical Chemistry Chemical Physics** 2016, *18*, 29045-29055
F. D. Steffen, R. K. O. Sigel, R. Börner
doi.org/10.1039/c6cp04277e (Front Cover)

IT and Data Science

Programming / OS	Python, Matlab, R, Unix
Molecular Modeling	Gromacs, Rosetta, PyMOL
Visualization	Adobe Photoshop / Illustrator

Languages

German (native) | **English** (full professional proficiency) | **French** (professional working proficiency) | **Spanish** (elementary proficiency)

Abbreviations

AdoCbl adenosyl cobalamin

AqCbl aquacobalamin

BVA burst variance analysis

Cas CRISPR-associated (protein)

CASP Critical Assessment of Methods for Protein Structure Prediction

CRISPR clustered regularly interspaced short palindromic repeats

DNA deoxyribonucleic acid

EBS exon binding site

EF-Tu elongation factor thermo unstable

EMCCD electron multiplying charge-coupled device

EPR electron paramagnetic resonance

FARFAR2 Fragment Assembly of RNA with Full-Atom Refinement

FCS fluorescence correlation spectroscopy

FMN flavin mononucleotide

FRET Förster resonance energy transfer

GTP guanosine triphosphate

IBS intron binding site

miRNA micro ribonucleic acid

mRNA messenger RNA

NMR nuclear magnetic resonance

PDA photon distribution analysis

PET photoinduced electron transfer

PIFE protein-induced fluorescence enhancement

ppGpp guanosine tetraphosphate

Prp8 pre-mRNA processing 8 (spliceosomal protein)

PRPP phosphoribosyl pyrophosphate

RBS ribosomal binding site

RIFE RNA-induced fluorescence enhancement

RISC RNA-induced silencing complex

RNA ribonucleic acid

RNAP RNA polymerase

RNP ribonucleoprotein

SAM S-adenosyl methionine

SAXS small-angle X-ray scattering

siRNA small interfering RNA

snRNA small nuclear RNA

SPAD single-photon avalanche diode

SRL sarcin-ricin loop

TCSPC time-correlated single photon counting

TIRF total internal reflection fluorescence

TOCSY total correlation spectroscopy

TPP thiamin pyrophosphate

tRNA transfer RNA

UTR untranslated region

WC Watson-Crick

XL-MS cross-linking with mass spectrometry

Bibliography

[1] R. P. Feynman, R. B. Leighton, M. Sands, *The Feynman lectures on physics*, New millennium ed., Basic Books, New York, **2013**.

[2] S. M. Berget, C. Moore, P. A. Sharp, *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 3171–3175.

[3] L. T. Chow, R. E. Gelinas, T. R. Broker, R. J. Roberts, *Cell* **1977**, *12*, 1–8.

[4] G. W. Beadle, E. L. Tatum, *Proc. Natl. Acad. Sci. U. S. A.* **1941**, *27*, 499–506.

[5] A. J. Matlin, F. Clark, C. W. J. Smith, *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 386–398.

[6] W. Gilbert, *Nature* **1978**, *271*, 501.

[7] A. J. Berk, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 801–805.

[8] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76–82.

[9] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833–836.

[10] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242.

[11] K. A. Jarrell, C. L. Peebles, R. C. Dietrich, S. L. Romiti, P. S. Perlman, *J. Biol. Chem.* **1988**, *263*, 3432–3439.

[12] C. L. Peebles, P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor, K. A. Jarrell, H.-L. Cheng, **1986**, *44*, 213–223.

[13] D. L. Daniels, W. J. Michels Jr, A. M. Pyle, *J. Mol. Biol.* **1996**, *256*, 31–49.

[14] K. Chin, A. M. Pyle, *RNA* **1995**, *1*, 391–406.

[15] T. A. Steitz, J. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 6498–6502.

[16] A. Pyle, *Science* **1993**, *261*, 709–714.

[17] F. H. C. Crick, *Symp. Soc. Exp. Biol.* **1958**, *12*, 138–163.

[18] F. H. C. Crick, *Nature* **1970**, *227*, 561–563.

[19] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147–157.

[20] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849–857.

[21] P. G. Higgs, N. Lehman, *Nat. Rev. Genet.* **2015**, *16*, 7–17.

[22] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.

[23] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 163.

[24] C. Hsiao, I.-C. Chou, C. D. Okafor, J. C. Bowman, E. B. O'Neill, S. S. Athavale, A. S. Petrov, N. V. Hud, R. M. Wartell, S. C. Harvey, L. D. Williams, *Nat. Chem.* **2013**, *5*, 525–528.

[25] T. R. Cech, J. A. Steitz, *Cell* **2014**, *157*, 77–94.

[26] W. P. Galej, N. Toor, A. J. Newman, K. Nagai, *Chem. Rev.* **2018**, *118*, 4156–4176.

[27] S. Zimmerly, C. Semper, *Mob. DNA* **2015**, *6*, 7.

[28] E. J. Sontheimer, P. M. Gordon, J. A. Piccirilli, *Genes Dev.* **1999**, *13*, 1729–1741.

[29] C. Zhao, A. M. Pyle, *Nat. Struct. Mol. Biol.* **2016**, *23*, 558–565.

[30] W. P. Galej, M. E. Wilkinson, S. M. Fica, C. Oubridge, A. J. Newman, K. Nagai, *Nature* **2016**, *537*, 197–201.

[31] S. Ray, J. R. Widom, N. G. Walter, *Chem. Rev.* **2018**, *118*, 4120–4155.

[32] R. C. Wilson, J. A. Doudna, *Annu. Rev. Biophys.* **2013**, *42*, 217–239.

[33] W. C. Winkler, S. Cohen-Chalamish, R. R. Breaker, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 15908–15913.

[34] A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, *Nature* **2006**, *441*, 1167–1171.

[35] W. C. Winkler, A. Nahvi, N. Sudarsan, J. E. Barrick, R. R. Breaker, *Nat. Struct. Biol.* **2003**, *10*, 701–707.

[36] R. K. Montange, R. T. Batey, *Nature* **2006**, *441*, 1172–1175.

[37] A. Serganov, L. Huang, D. J. Patel, *Nature* **2009**, *458*, 233–237.

[38] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, R. R. Breaker, *Chem. Biol.* **2002**, *9*, 1043.

[39] P. J. McCown, K. A. Corbino, S. Stav, M. E. Sherlock, R. R. Breaker, *RNA* **2017**, *23*, 995–1011.

[40] M. E. Sherlock, R. R. Breaker, *RNA* **2020**, *26*, 675–693.

[41] E. B. Greenlee, S. Stav, R. M. Atilho, K. I. Brewer, K. A. Harris, S. N. Malkowski, G. Mirihana Arachchilage, K. R. Perkins, M. E. Sherlock, R. R. Breaker, *RNA Biol.* **2018**, *15*, 377–390.

[42] M. E. Sherlock, N. Sudarsan, R. R. Breaker, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 6052–6057.

[43] M. E. Sherlock, N. Sudarsan, S. Stav, R. R. Breaker, *eLife* **2018**, *7*, e33908.

[44] A. Peselis, A. Serganov, *Nat. Chem. Biol.* **2018**, *14*, 887–894.

[45] J. E. Barrick, K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser, R. R. Breaker, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6421–6426.

[46] I. R. Price, A. Gaballa, F. Ding, J. D. Helmann, A. Ke, *Mol. Cell* **2015**, *57*, 1110–1123.

[47] K. C. Suddala, I. R. Price, S. S. Dandpat, M. Janeček, P. Kührová, J. Šponer, P. Banáš, A. Ke, N. G. Walter, *Nat. Commun.* **2019**, *10*, 4304.

[48] Z. Weinberg, C. E. Lünse, K. A. Corbino, T. D. Ames, J. W. Nelson, A. Roth, K. R. Perkins, M. E. Sherlock, R. R. Breaker, *Nucleic Acids Res.* **2017**, *45*, 10811–10823.

[49] R. R. Breaker, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003566.

[50] J. A. Howe, H. Wang, T. O. Fischmann, C. J. Balibar, L. Xiao, A. M. Galgoci, J. C. Malinverni, T. Mayhood, A. Villafania, A. Nahvi, N. Murgolo, C. M. Barbieri, P. A. Mann, D. Carr, E. Xia, P. Zuck, D. Riley, R. E. Painter, S. S. Walker, B. Sherborne, R. de Jesus, W. Pan, M. A. Plotkin, J. Wu, D. Rindgen, J. Cummings, C. G. Garlisi, R. Zhang, P. R. Sheth, C. J. Gill, H. Tang, T. Roemer, *Nature* **2015**, *526*, 672–677.

[51] K. F. Blount, R. R. Breaker, *Nat. Biotechnol.* **2006**, *24*, 1558–1564.

[52] J. L. Jenkins, J. Krucinska, R. M. McCarty, V. Bandarian, J. E. Wedekind, *J. Biol. Chem.* **2011**, *286*, 24626–24637.

[53] A. Ren, K. R. Rajashankar, D. J. Patel, *Nature* **2012**, *486*, 85–89.

[54] W. Winkler, A. Nahvi, R. R. Breaker, *Nature* **2002**, *419*, 952–956.

[55] A. S. Mironov, I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov, E. Nudler, *Cell* **2002**, *111*, 747–756.

[56] A. Serganov, E. Nudler, *Cell* **2013**, *152*, 17–24.

[57] M. Mandal, R. R. Breaker, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 451–463.

[58] I. Gusarov, E. Nudler, *Mol. Cell* **1999**, *3*, 495–504.

[59] W. S. Yarnell, J. W. Roberts, *Science* **1999**, *284*, 611–615.

[60] M. D. Lundrigan, W. Köster, R. J. Kadner, *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 1479–1483.

[61] S. Ravnum, D. I. Andersson, *Mol Microbiol* **1997**, *23*, 35–42.

[62] X. Nou, R. J. Kadner, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 7190–7195.

[63] A. A. Richter-Dahlfors, D. I. Andersson, *Mol Microbiol* **1992**, *6*, 743–749.

[64] X. Nou, R. J. Kadner, *J. Bacteriol.* **1998**, *180*, 6719–6728.

[65] A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, M. S. Gelfand, *RNA* **2003**, *9*, 1084–1097.

[66] A. Nahvi, J. E. Barrick, R. R. Breaker, *Nucleic Acids Res.* **2004**, *32*, 143–150.

[67] J. Zhang, M. W. Lau, A. R. Ferré-D'Amaré, *Biochemistry* **2010**, *49*, 9123–9131.

[68] J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, *Mol. Cell* **2005**, *18*, 49–60.

[69] A. Peselis, A. Serganov, *Biochim. Biophys. Acta* **2014**, *1839*, 908–918.

[70] A. Serganov, Y.-R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, D. J. Patel, *Chem. Biol.* **2004**, *11*, 1729–1741.

[71] A. Haller, M. F. Soulière, R. Micura, *Acc. Chem. Res.* **2011**, *44*, 1339–1348.

[72] S. D. Gilbert, R. P. Rambo, D. van Tyne, R. T. Batey, *Nat. Struct. Mol. Biol.* **2008**, *15*, 177–182.

[73] A. Haller, U. Rieder, M. Aigner, S. C. Blanchard, R. Micura, *Nat. Chem. Biol.* **2011**, *7*, 393–400.

[74] E. D. Holmstrom, J. T. Polaski, R. T. Batey, D. J. Nesbitt, *J. Am. Chem. Soc.* **2014**, *136*, 16832–16843.

[75] A. J. Rinaldi, P. E. Lund, M. R. Blanco, N. G. Walter, *Nat. Commun.* **2016**, *7*, 8976.

[76] G. A. Perdrizet, I. Artsimovitch, R. Furman, T. R. Sosnick, T. Pan, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3323–3328.

[77] R. Rieder, K. Lang, D. Graber, R. Micura, *ChemBioChem* **2007**, *8*, 896–902.

[78] U. Rieder, C. Kreutz, R. Micura, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10804–10809.

[79] C. Manz, A. Y. Kobitski, A. Samanta, B. G. Keller, A. Jäschke, G. U. Nienhaus, *Nat. Chem. Biol.* **2017**, *13*, 1172–1178.

[80] K. C. Suddala, J. Wang, Q. Hou, N. G. Walter, *J. Am. Chem. Soc.* **2015**, *137*, 14075–14083.

[81] A. Chauvier, F. Picard-Jean, J.-C. Berger-Dancause, L. Bastet, M. R. Naghdi, A. Dubé, P. Turcotte, J. Perreault, D. A. Lafontaine, *Nat. Commun.* **2017**, *8*, 13892.

[82] J.-F. Lemay, G. Desnoyers, S. Blouin, B. Heppell, L. Bastet, P. St-Pierre, E. Massé, D. A. Lafontaine, *PLoS Genet.* **2011**, *7*, e1001278.

[83] K. L. Frieda, S. M. Block, *Science* **2012**, *338*, 397–400.

[84] D. D. Shultis, M. D. Purdy, C. N. Banchs, M. C. Wiener, *Science* **2006**, *312*, 1396–1399.

[85] R. M. Jeter, B. M. Olivera, J. R. Roth, *J. Bacteriol.* **1984**, *159*, 206–213.

[86] J. E. Johnson, F. E. Reyes, J. T. Polaski, R. T. Batey, *Nature* **2012**, *492*, 133–137.

[87] A. Peselis, A. Serganov, *Nat. Struct. Mol. Biol.* **2012**, *19*, 1182–1184.

[88] Z. Weinberg, J. X. Wang, J. Bogue, J. Yang, K. Corbino, R. H. Moy, R. R. Breaker, *Genome Biol.* **2010**, *11*, R31.

[89] L. Jaeger, E. J. Verzemnieks, C. Geary, *Nucleic Acids Res.* **2009**, *37*, 215–230.

[90] A. Serganov, L. Huang, D. J. Patel, *Nature* **2008**, *455*, 1263–1267.

[91] S. Gallo, M. Oberhuber, R. K. O. Sigel, B. Kräutler, *ChemBioChem* **2008**, *9*, 1408–1414.

[92] D. E. Draper, *RNA* **2004**, *10*, 335–343.

[93] J. Lipfert, S. Doniach, R. Das, D. Herschlag, *Annu. Rev. Biochem.* **2014**, *83*, 813–841.

[94] J. C. Bowman, T. K. Lenz, N. V. Hud, L. D. Williams, *Curr. Opin. Struct. Biol.* **2012**, *22*, 262–272.

[95] S. A. Woodson, *Curr. Opin. Chem. Biol.* **2005**, *9*, 104–109.

[96] M. J. Cromie, Y. Shi, T. Latifi, E. A. Groisman, *Cell* **2006**, *125*, 71–84.

[97] C. E. Dann, C. A. Wakeman, C. L. Sieling, S. C. Baker, I. Irnov, W. C. Winkler, *Cell* **2007**, *130*, 878–892.

[98] K. Furukawa, A. Ramesh, Z. Zhou, Z. Weinberg, T. Vallery, W. C. Winkler, R. R. Breaker, *Mol. Cell* **2015**, *57*, 1088–1098.

[99] J. L. Baker, N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge, R. R. Breaker, *Science* **2012**, *335*, 233–235.

[100] V. Chandra, Z. Hannan, H. Xu, M. Mandal, *Nat. Chem. Biol.* **2017**, *13*, 194–201.

[101] H.-L. Sung, D. J. Nesbitt, *J. Phys. Chem. B* **2019**, *123*, 2005–2015.

[102] M. D. Brenner, M. S. Scanlan, M. K. Nahas, T. Ha, S. K. Silverman, *Biochemistry* **2010**, *49*, 1596–1605.

[103] A. Savinov, C. F. Perez, S. M. Block, *Biochim. Biophys. Acta* **2014**, *1839*, 1030–1045.

[104] N. J. Baird, N. Kulshina, A. R. Ferré D'Amaré, *RNA Biol.* **2014**, *7*, 328–332.

[105] D. K. Treiber, J. R. Williamson, *Curr. Opin. Struct. Biol.* **1999**, *9*, 339–345.

[106] M. A. Ditzler, D. Rueda, J. Mo, K. Håkansson, N. G. Walter, *Nucleic Acids Res.* **2008**, *36*, 7088–7099.

[107] S. V. Solomatin, M. Greenfeld, S. Chu, D. Herschlag, *Nature* **2010**, *463*, 681–684.

[108] S. A. Woodson, *Annu. Rev. Biophys.* **2010**, *39*, 61–77.

[109] L. R. Ganser, M. L. Kelly, D. Herschlag, H. M. Al-Hashimi, *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 474–489.

[110] N. A. Denesyuk, D. Thirumalai, *Nat. Chem.* **2015**, *7*, 793–801.

[111] R. K. O. Sigel, H. Sigel, *Acc. Chem. Res.* **2010**, *43*, 974–984.

[112] E. Freisinger, R. K. O. Sigel, *Coord. Chem. Rev.* **2007**, *251*, 1834–1851.

[113] D. J. Klein, P. B. Moore, T. A. Steitz, *RNA* **2004**, *10*, 1366–1379.

[114] M. Marcia, A. M. Pyle, *RNA* **2014**, *20*, 516–527.

[115] M. R. Stahley, P. L. Adams, J. Wang, S. A. Strobel, *J. Mol. Biol.* **2007**, *372*, 89–102.

[116] R. Yamagami, J. L. Bingaman, E. A. Frankel, P. C. Bevilacqua, *Nat. Commun.* **2018**, *9*, 2149.

[117] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. van Dien, J. D. Rabinowitz, *Nat. Chem. Biol.* **2009**, *5*, 593–599.

[118] A. M. P. Romani, *Arch. Biochem. Biophys.* **2011**, *512*, 1–23.

[119] D. Herschlag, *J. Biol. Chem.* **1995**, *270*, 20871–20874.

[120] N. B. Leontis, E. Westhof, *RNA* **2001**, *7*, 499–512.

[121] D. Herschlag, M. Khosla, Z. Tsuchihashi, R. L. Karpel, *EMBO J.* **1994**, *13*, 2913–2924.

[122] J. Pan, D. Thirumalai, S. A. Woodson, *J. Mol. Biol.* **1997**, *273*, 7–13.

[123] N. Bisaria, M. Greenfeld, C. Limouse, D. S. Pavlichin, H. Mabuchi, D. Herschlag, *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, E4956–65.

[124] Z. Miao, R. W. Adamiak, M. Antczak, R. T. Batey, A. J. Becka, M. Biesiada, M. J. Boniecki, J. M. Bujnicki, S.-J. Chen, C. Y. Cheng, F.-C. Chou, A. R. Ferré-D'Amaré, R. Das, W. K. Dawson, F. Ding, N. V. Dokholyan, S. Dunin-Horkawicz, C. Geniesse, K. Kappel, W. Kladwang, A. Krokhotin, G. E. Łach, F. Major, T. H. Mann, M. Magnus, K. Pachulska-Wieczorek, D. J. Patel, J. A. Piccirilli, M. Popenda, K. J. Purzycka, A. Ren, G. M. Rice, J. Santalucia, J. Sarzynska, M. Szachniuk, A. Tandon, J. J. Trausch, S. Tian, J. Wang, K. M. Weeks, B. Williams, Y. Xiao, X. Xu, D. Zhang, T. Zok, E. Westhof, *RNA* **2017**, *23*, 655–672.

[125] N. B. Leontis, A. Lescoute, E. Westhof, *Curr. Opin. Struct. Biol.* **2006**, *16*, 279–287.

[126] P. Brion, E. Westhof, *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113–137.

[127] B. D. Sattin, W. Zhao, K. Travers, S. Chu, D. Herschlag, *J. Am. Chem. Soc.* **2008**, *130*, 6085–6087.

[128] R. Behrouzi, J. H. Roh, D. Kilburn, R. M. Briber, S. A. Woodson, *Cell* **2012**, *149*, 348–357.

[129] S. Panja, B. Hua, D. Zegarra, T. Ha, S. A. Woodson, *Nat. Chem. Biol.* **2017**, *13*, 1109–1114.

[130] M. Steiner, K. S. Karunatilaka, R. K. O. Sigel, D. Rueda, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 13853–13858.

[131] N. F. Dupuis, E. D. Holmstrom, D. J. Nesbitt, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8464–8469.

[132] B. P. Paudel, E. Fiorini, R. Börner, R. K. O. Sigel, D. S. Rueda, *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, 11917–11922.

[133] E. D. Holmstrom, Z. Liu, D. Nettels, R. B. Best, B. Schuler, *Nat. Commun.* **2019**, *10*, 2453.

[134] S. C. Abeysirigunawardena, H. Kim, J. Lai, K. Ragunathan, M. C. Rappé, Z. Luthey-Schulten, T. Ha, S. A. Woodson, *Nat. Commun.* **2017**, *8*, 492.

[135] K. S. Karunatilaka, A. Solem, A. M. Pyle, D. Rueda, *Nature* **2010**, *467*, 935–939.

[136] D. H. Mathews, M. D. Disney, J. L. Childs, S. J. Schroeder, M. Zuker, D. H. Turner, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7287–7292.

[137] D. H. Mathews, D. H. Turner, *Curr. Opin. Struct. Biol.* **2006**, *16*, 270–278.

[138] N. Bisaria, M. Greenfeld, C. Limouse, H. Mabuchi, D. Herschlag, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E7688–E7696.

[139] D. Herschlag, S. Bonilla, N. Bisaria, *Cold Spring Harbor Perspect. Biol.* **2018**, *10*, 1–18.

[140] S. K. Denny, N. Bisaria, J. D. Yesselman, R. Das, D. Herschlag, W. J. Greenleaf, *Cell* **2018**, *174*, 377–390.e20.

[141] J. D. Yesselman, S. K. Denny, N. Bisaria, D. Herschlag, W. J. Greenleaf, R. Das, *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 16847–16855.

[142] J. S. Reuter, D. H. Mathews, *BMC Bioinf.* **2010**, *11*, 129.

[143] M. Zuker, *Nucleic Acids Res.* **2003**, *31*, 3406–3415.

[144] R. Lorenz, S. H. Bernhart, Höner Zu Siederdissen, Christian, H. Tafer, C. Flamm, P. F. Stadler, I. L. Hofacker, *Algorithms Mol. Biol.* **2011**, *6*, 26.

[145] P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 4899–4903.

[146] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, J. A. Doudna, *Science* **1996**, *273*, 1678–1685.

[147] A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, J. Moult, *Proteins* **2019**, *87*, 1011–1020.

[148] J. A. Cruz, M.-F. Blanchet, M. Boniecki, J. M. Bujnicki, S.-J. Chen, S. Cao, R. Das, F. Ding, N. V. Dokholyan, S. C. Flores, L. Huang, C. A. Lavender, V. Lisi, F. Major, K. Mikolajczak, D. J. Patel, A. Philips, T. Puton, J. Santalucia, F. Sijenyi, T. Hermann, K. Rother, M. Rother, A. Serganov, M. Skorupski, T. Soltysinski, P. Sripakdeevong, I. Tuszynska, K. M. Weeks, C. Waldsich, M. Wildauer, N. B. Leontis, E. Westhof, *RNA* **2012**, *18*, 610–625.

[149] Z. Miao, R. W. Adamiak, M.-F. Blanchet, M. Boniecki, J. M. Bujnicki, S.-J. Chen, C. Cheng, G. Chojnowski, F.-C. Chou, P. Cordero, J. A. Cruz, A. R. Ferré-D'Amaré, R. Das, F. Ding, N. V. Dokholyan, S. Dunin-Horkawicz, W. Kladwang, A. Krokhotin, G. Lach, M. Magnus, F. Major, T. H. Mann, B. Masquida, D. Matelska, M. Meyer, A. Peselis, M. Popenda, K. J. Purzycka, A. Serganov, J. Stasiewicz, M. Szachniuk, A. Tandon, S. Tian, J. Wang, Y. Xiao, X. Xu, J. Zhang, P. Zhao, T. Zok, E. Westhof, *RNA* **2015**, *21*, 1066–1084.

[150] Z. Miao, E. Westhof, *Annu. Rev. Biophys.* **2017**, *46*, 483–503.

[151] Z. Miao, R. W. Adamiak, M. Antczak, M. J. Boniecki, J. Bujnicki, S.-J. Chen, C. Y. Cheng, Y. Cheng, F.-C. Chou, R. Das, N. V. Dokholyan, F. Ding, C. Geniesse, Y. Jiang, A. Joshi, A. Krokhotin, M. Magnus, O. Mailhot, F. Major, T. H. Mann, P. Piatkowski, R. Pluta, M. Popenda, J. Sarzynska, L. Sun, M. Szachniuk, S. Tian, J. Wang, J. Wang, A. M. Watkins, J. Wiedemann, Y. Xiao, X. Xu, J. D. Yesselman, D. Zhang, Y. Zhang, Z. Zhang, C. Zhao, P. Zhao, Y. Zhou, T. Zok, A. Żyła, A. Ren, R. T. Batey, B. L. Golden, L. Huang, D. M. Lilley, Y. Liu, D. J. Patel, E. Westhof, *RNA* **2020**, *26*, 982–995.

[152] N. B. Leontis, E. Westhof, *Curr. Opin. Struct. Biol.* **2003**, *13*, 300–308.

[153] N. B. Leontis, E. Westhof, *J. Mol. Biol.* **1998**, *283*, 571–583.

[154] N. B. Leontis, J. Stombaugh, E. Westhof, *Nucleic Acids Res.* **2002**, *30*, 3497–3531.

[155] D. J. Klein, T. M. Schmeing, P. B. Moore, T. A. Steitz, *EMBO J.* **2001**, *20*, 4214–4221.

[156] A. M. Watkins, R. Rangan, R. Das, **2020**, 1–14.

[157] A. A. Szewczak, P. B. Moore, *J. Mol. Biol.* **1995**, *247*, 81–98.

[158] B. Wimberly, G. Varani, I. Tinoco, *Biochemistry* **1993**, *32*, 1078–1087.

[159] S. Lyskov, F.-C. Chou, S. Ó. Conchúir, B. S. Der, K. Drew, D. Kuroda, J. Xu, B. D. Weitzner, P. D. Renfrew, P. Sripakdeevong, B. Borgo, J. J. Havranek, B. Kuhlman, T. Kortemme, R. Bonneau, J. J. Gray, R. Das, *PLoS One* **2013**, *8*, e63906.

[160] M. Magnus, M. J. Boniecki, W. Dawson, J. M. Bujnicki, *Nucleic Acids Res.* **2016**, *44*, W315–9.

[161] C. Y. Cheng, F.-C. Chou, W. Kladwang, S. Tian, P. Cordero, R. Das, *eLife* **2015**, *4*, e07600.

[162] K. Kappel, K. Zhang, Z. Su, A. M. Watkins, W. Kladwang, S. Li, G. Pintilie, V. V. Topkar, R. Rangan, I. N. Zheludev, J. D. Yesselman, W. Chiu, R. Das, *Nat. Methods* **2020**, *17*, 699–707.

[163] A. M. Watkins, C. Geniesse, W. Kladwang, P. Zakrevsky, L. Jaeger, R. Das, *Sci. Adv.* **2018**, *4*, eaar5316.

[164] D. Shortle, K. T. Simons, D. Baker, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 11158–11162.

[165] M. Kasha, *Discuss. Faraday Soc.* **1950**, *9*, 14.

[166] J. R. Lakowicz, *Principles of fluorescence spectroscopy*, 3rd ed., Springer, New York, **2006**.

[167] M. Levitus, S. Ranjit, *Q. Rev. Biophys.* **2011**, *44*, 123–151.

[168] E. M. S. Stennett, M. A. Ciuba, S. Lin, M. Levitus, *J. Phys. Chem. Lett.* **2015**, *6*, 1819–1823.

[169] E. Ploetz, E. Lerner, F. Husada, M. Roelfs, S. Chung, J. Hohlbein, S. Weiss, T. Cordes, *Sci. Rep.* **2016**, *6*, 33257.

[170] E. Lerner, E. Ploetz, J. Hohlbein, T. Cordes, S. Weiss, *J. Phys. Chem. B* **2016**, *120*, 6401–6410.

[171] F. D. Steffen, R. K. O. Sigel, R. Börner, *Phys. Chem. Chem. Phys.* **2016**, *18*, 29045–29055.

[172] H. Hwang, H. Kim, S. Myong, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 7414–7418.

[173] B. Valeur, *Molecular fluorescence: Principles and applications*, Wiley-VCH, Weinheim and New York, **2002**.

[174] R. F. Kubin, A. N. Fletcher, *J. Lumin.* **1982**, *27*, 455–462.

[175] R. Börner, D. Kowerko, H. Miserachs-Guiset, M. F. Schaffer, R. K. O. Sigel, *Coord. Chem. Rev.* **2016**, *327-328*, 123–142.

[176] M. Sustarsic, A. N. Kapanidis, *Curr. Opin. Struct. Biol.* **2015**, *34*, 52–59.

[177] F. Zosel, D. Mercadante, D. Nettels, B. Schuler, *Nat. Commun.* **2018**, *9*, 3332.

[178] F. D. Steffen, M. Khier, D. Kowerko, R. A. Cunha, R. Börner, R. K. O. Sigel, *Nat. Commun.* **2020**, *11*, 104.

[179] N. Vušurović, R. B. Altman, D. S. Terry, R. Micura, S. C. Blanchard, *J. Am. Chem. Soc.* **2017**, *139*, 8186–8193.

[180] M. F. Juette, D. S. Terry, M. R. Wasserman, R. B. Altman, Z. Zhou, H. Zhao, S. C. Blanchard, *Nat. Methods* **2016**, *13*, 341–344.

[181] S. Farooq, J. Hohlbein, *Phys. Chem. Chem. Phys.* **2015**, *17*, 27862–27872.

[182] S. Schmid, M. Götz, T. Hugel, *Biophys. J.* **2016**, *111*, 1375–1384.

[183] H. Kim, S. C. Abeysirigunawarden, K. Chen, M. Mayerle, K. Ragunathan, Z. Luthey-Schulten, T. Ha, S. A. Woodson, *Nature* **2014**, *506*, 334–338.

[184] T. Torres, M. Levitus, *J. Phys. Chem. B* **2007**, *111*, 7392–7400.

[185] J. Widengren, E. Schweinberger, S. Berger, C. A. M. Seidel, *J. Phys. Chem. A* **2001**, *105*, 6851–6866.

[186] B. Schuler, *J. Chem. Phys.* **2018**, *149*, 010901.

[187] J. P. Torella, S. J. Holden, Y. Santoso, J. Hohlbein, A. N. Kapanidis, *Biophys. J.* **2011**, *100*, 1568–1577.

[188] S. Kalinin, A. Valeri, M. Antonik, S. Felekyan, C. A. M. Seidel, *J. Phys. Chem. B* **2010**, *114*, 7983–7995.

[189] T. E. Tomov, R. Tsukanov, R. Masoud, M. Liber, N. Plavner, E. Nir, *Biophys. J.* **2012**, *102*, 1163–1173.

[190] N. K. Lee, A. N. Kapanidis, Y. Wang, X. Michalet, J. Mukhopadhyay, R. H. Ebright, S. Weiss, *Biophys. J.* **2005**, *88*, 2939–2953.

[191] B. Hellenkamp, S. Schmid, O. Doroshenko, O. Opanasyuk, R. Kühnemuth, S. Rezaei Adariani, B. Ambrose, M. Aznauryan, A. Barth, V. Birkedal, M. E. Bowen, H. Chen, T. Cordes, T. Eilert, C. Fijen, C. Gebhardt, M. Götz, G. Gouridis, E. Gratton, T. Ha, P. Hao, C. A. Hanke, A. Hartmann, J. Hendrix, L. L. Hildebrandt, V. Hirschfeld, J. Hohlbein, B. Hua, C. G. Hübner, E. Kallis, A. N. Kapanidis, J.-Y. Kim, G. Krainer, D. C. Lamb, N. K. Lee, E. A. Lemke, B. Levesque, M. Levitus, J. J. McCann, N. Naredi-Rainer, D. Nettels, T. Ngo, R. Qiu, N. C. Robb, C.

Röcker, H. Sanabria, M. Schlierf, T. Schröder, B. Schuler, H. Seidel, L. Streit, J. Thurn, P. Tinnefeld, S. Tyagi, N. Vandenberk, A. M. Vera, K. R. Weninger, B. Wünsch, I. S. Yanez-Orozco, J. Michaelis, C. A. M. Seidel, T. D. Craggs, T. Hugel, *Nat. Methods* **2018**, *15*, 669–676.

[192] A. Hoffmann, D. Nettels, J. Clark, A. Borgia, S. E. Radford, J. Clarke, B. Schuler, *Phys. Chem. Chem. Phys.* **2011**, *13*, 1857–1871.

[193] I. V. Gopich, D. Nettels, B. Schuler, A. Szabo, *J. Chem. Phys.* **2009**, *131*, 095102.

[194] D. Nettels, I. V. Gopich, A. Hoffmann, B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 2655–2660.

[195] J. Kubelka, J. Hofrichter, W. A. Eaton, *Curr. Opin. Struct. Biol.* **2004**, *14*, 76–88.

[196] B. Schuler, *J. Nanobiotechnology* **2013**, *11*, S2.

[197] J. C. Morse, D. Girodat, B. J. Burnett, M. Holm, R. B. Altman, K. Y. Sanbonmatsu, H.-J. Wieden, S. C. Blanchard, *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 3610–3620.

[198] T. Förster, *Ann. Phys.* **1948**, *437*, 55–75.

[199] T. Eilert, E. Kallis, J. Nagy, C. Röcker, J. Michaelis, *J. Phys. Chem. B* **2018**, *122*, 11677–11694.

[200] R. E. Dale, J. Eisinger, W. E. Blumberg, *Biophys. J.* **1979**, *26*, 161–193.

[201] M. Hoefling, N. Lima, D. Haenni, C. A. M. Seidel, B. Schuler, H. Grubmüller, *PLoS One* **2011**, *6*, e19791.

[202] R. B. Best, H. Hofmann, D. Nettels, B. Schuler, *Biophys. J.* **2015**, *108*, 2721–2731.

[203] G. F. Schröder, U. Alexiev, H. Grubmüller, *Biophys. J.* **2005**, *89*, 3757–3770.

[204] M. Dimura, T. O. Peulen, C. A. Hanke, A. Prakash, H. Gohlke, C. A. M. Seidel, *Curr. Opin. Struct. Biol.* **2016**, *40*, 163–185.

[205] T.-O. Peulen, O. Opanasyuk, C. A. M. Seidel, *J. Phys. Chem. B* **2017**, *121*, 8211–8241.

[206] R. S. Knox, H. van Amerongen, *J. Phys. Chem. B* **2002**, *106*, 5289–5293.

[207] J. J. McCann, U. B. Choi, L. Zheng, K. Weninger, M. E. Bowen, *Biophys. J.* **2010**, *99*, 961–970.

[208] A. Hoffmann, A. Kane, D. Nettels, D. E. Hertzog, P. Baumgärtel, J. Lengefeld, G. Reichardt, D. A. Horsley, R. Seckler, O. Bakajin, B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 105–110.

[209] D. Nettels, S. Müller-Späth, F. Küster, H. Hofmann, D. Haenni, S. Rüegger, L. Reymond, A. Hoffmann, J. Kubelka, B. Heinz, K. Gast, R. B. Best, B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 20740–20745.

[210] S. Müller-Späth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rüegger, L. Reymond, D. Nettels, B. Schuler, *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 14609–14614.

[211] B. Schuler, E. A. Lipman, P. J. Steinbach, M. Kumke, W. A. Eaton, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 2754–2759.

[212] W. Zheng, G. H. Zerze, A. Borgia, J. Mittal, B. Schuler, R. B. Best, *J. Chem. Phys.* **2018**, *148*, 123329.

[213] B. Schuler, A. Soranno, H. Hofmann, D. Nettels, *Annu. Rev. Biophys.* **2016**, *45*, 207–231.

[214] E. D. Holmstrom, A. Holla, W. Zheng, D. Nettels, R. B. Best, B. Schuler, *Methods Enzymol.* **2018**, *611*, 287–325.

[215] L. Stryer, R. P. Haugland, *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58*, 719–726.

[216] T. Ha, T. Enderle, D. F. Ogletree, D. S. Chemla, P. R. Selvin, S. Weiss, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6264–6268.

[217] A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels, B. B. Kragelund, R. B. Best, B. Schuler, *Nature* **2018**, *555*, 61–66.

[218] E. Lerner, B. Ambrose, A. Barth, V. Birkedal, S. C. Blanchard, R. Börner, T. Cordes, T. D. Craggs, T. Ha, G. Haran, T. Hugel, A. Ingargiola, A. Kapanidis, D. C. Lamb, T. Laurence, N. K. Lee, E. A. Lemke, E. Margeat, J. Michaelis, X. Michalet, D. Nettels, T.-O. Peulen, B. Schuler, C. A. M. Seidel, H. So-leimaninejad, S. Weiss, **2020**.

[219] W. R. Algar, N. Hildebrandt, S. S. Vogel, I. L. Medintz, *Nat. Methods* **2019**, *16*, 815–829.

[220] B. Hellenkamp, P. Wortmann, F. Kandzia, M. Zacharias, T. Hugel, *Nat. Methods* **2017**, *14*, 174–180.

[221] M. P. Rout, A. Sali, *Cell* **2019**, *177*, 1384–1403.

[222] H. Sanabria, D. Rodnin, K. Hemmen, T.-O. Peulen, S. Felekyan, M. R. Fleissner, M. Dimura, F. Koberling, R. Kühnemuth, W. Hubbell, H. Gohlke, C. A. M. Seidel, *Nat. Commun.* **2020**, *11*, 1231.

[223] G. Dorn, A. Leitner, J. Boudet, S. Campagne, C. von Schroetter, A. Moursy, R. Aebersold, F. H.-T. Allain, *Nat. Methods* **2017**, *14*, 487–490.

[224] O. Klykov, C. van der Zwaan, A. J. R. Heck, A. B. Meijer, R. A. Scheltema, **2020**, *117*, 1976–1987.

[225] S. Bonilla, C. Limouse, N. Bisaria, M. Gebala, H. Mabuchi, D. Herschlag, *J. Am. Chem. Soc.* **2017**, *139*, 18576–18589.

[226] J. L. Fiore, E. D. Holmstrom, D. J. Nesbitt, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 2902–2907.

[227] J.-Y. Roignant, M. Soller, *Trends Genet.* **2017**, *33*, 380–390.

[228] C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, *Angew. Chem. Int. Ed.* **2005**, *44*, 7342–7372.

[229] J. A. Prescher, C. R. Bertozzi, *Nat. Chem. Biol.* **2005**, *1*, 13–21.

[230] D. Schulz, A. Rentmeister, *ChemBioChem* **2014**, *15*, 2342–2347.

[231] S. Saha, A. P. Jagtap, S. T. Sigurdsson, *Methods Enzymol.* **2015**, *563*, 397–414.

[232] S. Solomatin, D. Herschlag, *Methods Enzymol.* **2009**, *469*, 47–68.

[233] S. Seidu-Larry, B. Krieg, M. Hirsch, M. Helm, O. Domingo, *Chem. Commun.* **2012**, *48*, 11014–11016.

[234] S. H. Weisbrod, A. Marx, *Chem. Commun.* **2008**, 5675–5685.

[235] M. J. Moore, P. A. Sharp, *Science* **1992**, *256*, 992–997.

[236] J. M. Esquiaqui, E. M. Sherman, J.-D. Ye, G. E. Fanucci, *Methods Enzymol.* **2014**, *549*, 287–311.

[237] K. Lang, R. Micura, *Nat. Protoc.* **2008**, *3*, 1457–1466.

[238] D. A. Baum, S. K. Silverman, *Angew. Chem. Int. Ed.* **2007**, *46*, 3502–3504.

[239] L. Büttner, F. Javadi-Zarnaghi, C. Höbartner, *J. Am. Chem. Soc.* **2014**, *136*, 8131–8137.

[240] L. Anhäuser, A. Rentmeister, *Curr. Opin. Biotechnol.* **2017**, *48*, 69–76.

[241] A. Plotnikova, A. Osipenko, V. Masevičius, G. Vilkaitis, S. Klimašauskas, *J. Am. Chem. Soc.* **2014**, *136*, 13550–13553.

[242] J. M. Holstein, L. Anhäuser, A. Rentmeister, *Angew. Chem. Int. Ed.* **2016**, *55*, 10899–10903.

[243] M. Tomkuviene, B. Clouet-d'Orval, I. Cerniauskas, E. Weinhold, S. Klimasauskas, *Nucleic Acids Res.* **2012**, *40*, 6765–6773.

[244] J. S. Paige, K. Y. Wu, S. R. Jaffrey, *Science* **2011**, *333*, 642–646.

[245] M. Zhao, F. D. Steffen, R. Börner, M. F. Schaffer, R. K. O. Sigel, E. Freisinger, *Nucleic Acids Res.* **2018**, *46*, e13.

[246] A. G. Schmitz, S. Zelger-Paulus, G. Gasser, R. K. O. Sigel, *ChemBioChem* **2015**, *16*, 1302–1306.

[247] G. J. Smith, T. R. Sosnick, N. F. Scherer, T. Pan, *RNA* **2005**, *11*, 234–239.

[248] D. Egloff, I. A. Oleinich, E. Freisinger, *ACS Chem. Biol.* **2015**, *10*, 547–553.

[249] K. Jahn, E. M. Olsen, M. M. Nielsen, T. Tørring, R. Mohammad Zadegan, E. S. Andersen, K. V. Gothelf, J. Kjems, *Bioconjugate Chem.* **2011**, *22*, 95–100.

[250] K. Onizuka, Y. Taniguchi, S. Sasaki, *Bioconjugate Chem.* **2009**, *20*, 799–803.

[251] D. Egloff, I. A. Oleinich, M. Zhao, S. L. B. König, R. K. O. Sigel, E. Freisinger, *ACS Chem. Biol.* **2016**, *11*, 2558–2567.

[252] B. M. Fuchs, F. O. Glockner, J. Wulf, R. Amann, *Appl. Environ. Microbiol.* **2000**, *66*, 3603–3607.

[253] D. Proudnikov, A. Mirzabekov, *Nucleic Acids Res.* **1996**, *24*, 4535–4542.

[254] P. Z. Qin, A. M. Pyle, *Methods* **1999**, *18*, 60–70.

[255] B. Schuler, H. Hofmann, *Curr. Opin. Struct. Biol.* **2013**, *23*, 36–47.

[256] M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, R. R. Breaker, **2004**, *306*, 275–279.

[257] T. Ha, *Methods* **2001**, *25*, 78–86.

[258] S. Sindbert, S. Kalinin, H. Nguyen, A. Kienzler, L. Clima, W. Bannwarth, B. Appel, S. Müller, C. A. M. Seidel, *J. Am. Chem. Soc.* **2011**, *133*, 2463–2480.

[259] A. Iqbal, S. Arslan, B. Okumus, T. J. Wilson, G. Giraud, D. G. Norman, T. Ha, Lilley, David M J, *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 11176–11181.

[260] M. E. Sanborn, B. K. Connolly, K. Gurunathan, M. Levitus, *J. Phys. Chem. B* **2007**, *111*, 11064–11074.

[261] R. Roy, S. Hohng, T. Ha, *Nat. Methods* **2008**, *5*, 507–516.

[262] A. N. Kapanidis, N. K. Lee, T. A. Laurence, S. Doose, E. Margeat, S. Weiss, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 8936–8941.

[263] L. Kronberg, R. Sjöholm, S. Karlsson, *Chem. Res. Toxicol.* **1992**, *5*, 852–855.

[264] N. J. Agard, J. A. Prescher, C. R. Bertozzi, *J. Am. Chem. Soc.* **2004**, *126*, 15046–15047.

[265] E. Saxon, C. R. Bertozzi, *Science* **2000**, *287*, 2007–2010.

[266] J. Schoch, S. Ameta, A. Jäschke, *Chem. Commun.* **2011**, *47*, 12536–12537.

[267] A. M. Pyle, *Annu. Rev. Biophys.* **2016**, *45*, 183–205.

[268] A. Jacquier, F. Michel, *Cell* **1987**, *50*, 17–29.

[269] P. L. Adams, M. R. Stahley, A. B. Kosek, J. Wang, S. A. Strobel, *Nature* **2004**, *430*, 45–50.

[270] R. W. Davies, R. B. Waring, J. A. Ray, T. A. Brown, C. Scazzocchio, *Nature* **1982**, *300*, 719–724.

[271] L. J. Su, P. Z. Qin, W. J. Michels, A. M. Pyle, *J. Mol. Biol.* **2001**, *306*, 655–668.

[272] N. Toor, K. S. Keating, S. D. Taylor, A. M. Pyle, *Science* **2008**, *320*, 77–82.

[273] R. T. Chan, A. R. Robart, K. R. Rajashankar, A. M. Pyle, N. Toor, *Nat. Struct. Mol. Biol.* **2012**, *19*, 555–557.

[274] A. R. Robart, R. T. Chan, J. K. Peters, K. R. Rajashankar, N. Toor, *Nature* **2014**, *514*, 193–197.

[275] R. K. O. Sigel, A. Vaidya, A. M. Pyle, *Nat. Struct. Biol.* **2000**, *7*, 1111–1116.

[276] P. M. Gordon, J. A. Piccirilli, *Nat. Struct. Biol.* **2001**, *8*, 893–898.

[277] P. L. Deininger, M. A. Batzer, *Genome Res.* **2002**, *12*, 1455–1465.

[278] Q. Xiang, P. Z. Qin, W. J. Michels, K. Freeland, A. M. Pyle, *Biochemistry* **1998**, *37*, 3839–3849.

[279] S. Somarowthu, M. Legiewicz, K. S. Keating, A. M. Pyle, *Nucleic Acids Res.* **2014**, *42*, 1947–1958.

[280] D. Kowerko, S. L. B. König, M. Skilandat, D. Kruschel, M. C. A. S. Hadzic, L. Cardo, R. K. O. Sigel, *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 3403–3408.

[281] C. Hyeon, J. Lee, J. Yoon, S. Hohng, D. Thirumalai, *Nat. Chem.* **2012**, *4*, 907–914.

[282] G. Bokinsky, X. Zhuang, *Acc. Chem. Res.* **2005**, *38*, 566–573.

[283] R. K. Montange, R. T. Batey, *Annu. Rev. Biophys.* **2008**, *37*, 117–133.

[284] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 905–920.

[285] D. W. Staple, S. E. Butcher, *PLoS Biol.* **2005**, *3*, e213.

[286] A. M. Lambowitz, S. Zimmerly, *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, a003616.

[287] R. Eskes, J. Yang, A. M. Lambowitz, P. S. Perlman, *Cell* **1997**, *88*, 865–874.

[288] J. Yang, S. Zimmerly, P. S. Perlman, A. M. Lambowitz, *Nature* **1996**, *381*, 332–335.

[289] D. B. Haack, X. Yan, C. Zhang, J. Hingey, D. Lyumkis, T. S. Baker, N. Toor, *Cell* **2019**, *178*, 612–623.

[290] G. Qu, P. S. Kaushal, J. Wang, H. Shigematsu, C. L. Piazza, R. K. Agrawal, M. Belfort, H.-W. Wang, *Nat. Struct. Mol. Biol.* **2016**, *23*, 549–557.

[291] D. Kruschel, M. Skilandat, R. K. O. Sigel, *RNA* **2014**, *20*, 295–307.

[292] M. Skilandat, R. K. O. Sigel, *J. Biol. Chem.* **2014**, *289*, 20650–20663.

[293] E. A. Lesnik, S. M. Freier, *Biochemistry* **1995**, *34*, 10807–10815.

[294] K. B. Hall, L. W. McLaughlin, *Biochemistry* **1991**, *30*, 10606–10613.

[295] B. Rauzan, E. McMichael, R. Cave, L. R. Sevcik, K. Ostrosky, E. Whitman, R. Stegemann, A. L. Sinclair, M. J. Serra, A. A. Deckert, *Biochemistry* **2013**, *52*, 765–772.

[296] V. K. Misra, D. E. Draper, *J. Mol. Biol.* **2002**, *317*, 507–521.

[297] G. Bokinsky, D. Rueda, V. K. Misra, M. M. Rhodes, A. Gordus, H. P. Babcock, N. G. Walter, X. Zhuang, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9302–9307.

[298] L. E. Bartley, X. Zhuang, R. Das, S. Chu, D. Herschlag, *J. Mol. Biol.* **2003**, *328*, 1011–1026.

[299] R. A. Cunha, G. Bussi, *RNA* **2017**, *23*, 628–638.

[300] S. Wang, E. T. Kool, *Biochemistry* **1995**, *34*, 4125–4132.

[301] Batey, Rambo, Doudna, *Angew. Chem. Int. Ed.* **1999**, *38*, 2326–2343.

[302] W. Saenger, *Principles of nucleic acid structure*, Springer, New York, **1984**.

[303] M. Karplus, *J. Chem. Phys.* **1959**, *30*, 11–15.

[304] M. C. Clay, L. R. Ganser, D. K. Merriman, H. M. Al-Hashimi, *Nucleic Acids Res.* **2017**, *45*, e134.

[305] W. K. Olson, J. L. Sussman, *J. Am. Chem. Soc.* **1982**, *104*, 270–278.

[306] E. Duchardt, H. Schwalbe, *J. Biomol. NMR* **2005**, *32*, 295–308.

[307] J. I. Gyi, A. N. Lane, G. L. Conn, T. Brown, *Biochemistry* **1998**, *37*, 73–80.

[308] M. Salazar, O. Y. Fedoroff, J. M. Miller, N. S. Ribeiro, B. R. Reid, *Biochemistry* **1993**, *32*, 4207–4215.

[309] J. A. Cowan, *J. Am. Chem. Soc.* **1991**, *113*, 675–676.

[310] H.-R. Huang, C. E. Rowe, S. Mohr, Y. Jiang, A. M. Lambowitz, P. S. Perlman, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 163–168.

[311] B. Séraphin, M. Simon, A. Boulet, G. Faye, *Nature* **1989**, *337*, 84–87.

[312] H. Wank, J. SanFilippo, R. N. Singh, M. Matsuura, A. M. Lambowitz, *Mol. Cell* **1999**, *4*, 239–250.

[313] Y. Aizawa, Q. Xiang, A. M. Lambowitz, A. M. Pyle, *Mol. Cell* **2003**, *11*, 795–805.

[314] F. D. Steffen, R. Börner, E. Freisinger, R. K. O. Sigel, *Chimia* **2019**, *73*, 257–261.

[315] Y. Liu, E. Holmstrom, J. Zhang, P. Yu, J. Wang, M. A. Dyba, D. Chen, J. Ying, S. Lockett, D. J. Nesbitt, A. R. Ferré-D'Amaré, R. Sousa, J. R. Stagno, Y.-X. Wang, *Nature* **2015**, *522*, 368–372.

[316] N. Salim, R. Lamichhane, R. Zhao, T. Banerjee, J. Philip, D. Rueda, A. L. Feig, *Biophys. J.* **2012**, *102*, 1097–1107.

[317] J. Šponer, G. Bussi, M. Krepl, P. Banáš, S. Bottaro, R. A. Cunha, A. Gil-Ley, G. Pinamonti, S. Poblete, P. Jurečka, N. G. Walter, M. Otyepka, *Chem. Rev.* **2018**, *118*, 4177–4338.

[318] S. Schmid, T. Hugel, *J. Chem. Phys.* **2018**, *148*, 123312.

[319] M. C. A. S. Hadzic, R. Börner, S. L. B. König, D. Kowerko, R. K. O. Sigel, *J. Phys. Chem. B* **2018**, *122*, 6134–6147.

[320] S. L. B. König, M. C. A. S. Hadzic, E. Fiorini, R. Börner, D. Kowerko, W. U. Blanckenhorn, R. K. O. Sigel, *PLoS One* **2013**, *8*, e84157.

[321] H. Eyring, *J. Chem. Phys.* **1935**, *3*, 107–115.

[322] R. Börner, D. Kowerko, M. C. A. S. Hadzic, König, Sebastian L B, M. Ritter, R. K. O. Sigel, *PLoS One* **2018**, *13*, e0195277.

[323] E. Jurrus, D. Engel, K. Star, K. Monson, J. Brandi, L. E. Felberg, D. H. Brookes, L. Wilson, J. Chen, K. Liles, M. Chun, P. Li, D. W. Gohara, T. Dolinsky, R. Konecny, D. R. Koes, J. E. Nielsen, T. Head-Gordon, W. Geng, R. Krasny, G.-W. Wei, M. J. Holst, J. A. McCammon, N. A. Baker, *Protein Sci.* **2018**, *27*, 112–128.

[324] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, N. A. Baker, *Nucleic Acids Res.* **2007**, *35*, W522–5.

[325] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX* **2015**, *1-2*, 19–25.

[326] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[327] A. Pérez, I. Marchán, D. Svozil, J. Sponer, T. E. Cheatham, C. A. Laughton, M. Orozco, *Biophys. J.* **2007**, *92*, 3817–3829.

[328] P. Banáš, D. Hollas, M. Zgarbová, P. Jurečka, M. Orozco, T. E. Cheatham, J. Šponer, M. Otyepka, *J. Chem. Theory Comput.* **2010**, *6*, 3836–3849.

[329] M. Zgarbová, M. Otyepka, J. Sponer, A. Mládek, P. Banáš, T. E. Cheatham, P. Jurečka, *J. Chem. Theory Comput.* **2011**, *7*, 2886–2902.

[330] I. S. Joung, T. E. Cheatham, *J. Phys. Chem. B* **2008**, *112*, 9020–9041.

[331] P. Li, B. P. Roberts, D. K. Chakravorty, K. M. Merz, *J. Chem. Theory Comput.* **2013**, *9*, 2733–2748.

[332] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.* **2014**, *185*, 604–613.

[333] M. Huang, T. J. Giese, T.-S. Lee, D. M. York, *J. Chem. Theory Comput.* **2014**, *10*, 1538–1545.

[334] C. Altona, M. Sundaralingam, *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.

[335] W. Lee, M. Tonelli, J. L. Markley, *Bioinformatics* **2015**, *31*, 1325–1327.

[336] D. Kruschel, R. K. O. Sigel, *J. Inorg. Biochem.* **2008**, *102*, 2147–2154.

[337] D. L. Floyd, S. C. Harrison, A. M. van Oijen, *Biophys. J.* **2010**, *99*, 360–366.

[338] C. A. Coelho, J. T. Mexia, *Sankhya Ser. A* **2007**, *69*, 221–255.

[339] V. B. Chu, Y. Bai, J. Lipfert, D. Herschlag, S. Doniach, *Curr. Opin. Chem. Biol.* **2008**, *12*, 619–625.

[340] S. Malhotra, S. Träger, M. Dal Peraro, M. Topf, *Curr. Opin. Struct. Biol.* **2019**, *58*, 105–114.

[341] G. Jeschke, *Annu. Rev. Phys. Chem.* **2012**, *63*, 419–446.

[342] A. Muschielok, J. Andrecka, A. Jawhari, F. Bruckner, P. Cramer, J. Michaelis, *Nat. Methods* **2008**, *5*, 965–971.

[343] M. Hoefling, H. Grubmüller, *Comput. Phys. Commun.* **2013**, *184*, 841–852.

[344] P. K. Choudhary, R. K. O. Sigel, *RNA* **2014**, *20*, 36–45.

[345] P. K. Choudhary, S. Gallo, R. K. O. Sigel, *Front. Chem.* **2017**, *5*, 10.

[346] R. S. Gregorian, D. M. Crothers, *J. Mol. Biol.* **1995**, *248*, 968–984.

[347] D. Donghi, M. Pechlaner, C. Finazzo, B. Knobloch, R. K. O. Sigel, *Nucleic Acids Res.* **2013**, *41*, 2489–2504.

[348] H. N. Motlagh, J. O. Wrabl, J. Li, V. J. Hilser, *Nature* **2014**, *508*, 331–339.

[349] A. Peselis, A. Gao, A. Serganov, *Biochimie* **2015**, *117*, 100–109.

[350] Q. Vicens, E. Mondragón, R. T. Batey, *Nucleic Acids Res.* **2011**, *39*, 8586–8598.

[351] S. Gallo, M. Furler, R. K. O. Sigel, *Chimia* **2005**, *59*, 812–816.

[352] M. Zhao, R. Börner, R. K. O. Sigel, E. Freisinger, *Methods Mol. Biol.* **2020**, *2106*, 253–270.

[353] B. K. Müller, E. Zaychikov, C. Brauchle, D. C. Lamb, *Biophys. J.* **2005**, *89*, 3508–3522.

[354] W. Schrimpf, A. Barth, J. Hendrix, D. C. Lamb, *Biophys. J.* **2018**, *114*, 1518–1528.

[355] E. Nir, X. Michalet, K. M. Hamadani, T. A. Laurence, D. Neuhauser, Y. Kovchegov, S. Weiss, *J. Phys. Chem. B* **2006**, *110*, 22103–22124.

[356] I. Kalvari, J. Argasinska, N. Quinones-Olvera, E. P. Nawrocki, E. Rivas, S. R. Eddy, A. Bateman, R. D. Finn, A. I. Petrov, *Nucleic Acids Res.* **2018**, *46*, D335–D342.

[357] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, *Cell* **2003**, *113*, 577–586.

[358] J. T. Polaski, S. M. Webster, J. E. Johnson, R. T. Batey, *J. Biol. Chem.* **2017**, *292*, 11650–11658.

[359] E. P. Nawrocki, S. R. Eddy, *Bioinformatics* **2013**, *29*, 2933–2935.

[360] A. M. Watkins, R. Rangan, R. Das, *Methods Enzymol.* **2019**, *623*, 177–207.

[361] N. B. Leontis, C. L. Zirbel in *RNA 3D Structure Analysis and Prediction*, (Eds.: N. Leontis, E. Westhof), Nucleic Acids and Molecular Biology, Springer Berlin Heidelberg, Berlin, Heidelberg, **2012**, pp. 281–298.

[362] S. Kalinin, T. Peulen, S. Sindbert, P. J. Rothwell, S. Berger, T. Restle, R. S. Goody, H. Gohlke, C. A. M. Seidel, *Nat. Methods* **2012**, *9*, 1218–1225.

[363] T. Graen, M. Hoefling, H. Grubmüller, *J. Chem. Theory Comput.* **2014**, *10*, 5505–5512.

[364] Sousa da Silva, Alan W, W. F. Vranken, *BMC Res. Notes* **2012**, *5*, 367.

[365] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 09, Revision D.01, Wallingford, **2016**.

[366] P. Cieplak, W. D. Cornell, C. Bayly, P. A. Kollman, *J. Comput. Chem.* **1995**, *16*, 1357–1377.

[367] D. A. Case, K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, V. W. D. Cruzeiro, T. A. Darden, R. E. Duke, G. Giambasu, M. K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S. A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T. S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K. M. Merz, Y. Miao, O.

Mikhailovskii, Monard, G. Nguyen, H., A. Onufriev, F. Pan, S. Pantano, R. Qi, D. R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. L. Simmerling, N. R. Skrynnikov, J. Smith, J. Swails, R. C. Walker, J. Wang, L. Wilson, R. M. Wolf, X. Wu, Y. Xiong, Y. Xue, D. M. York, P. A. Kollman, AMBER 2020, San Francisco, **2020**.

[368] M. Zgarbová, F. J. Luque, J. Sponer, T. E. Cheatham, M. Otyepka, P. Jurečka, *J. Chem. Theory Comput.* **2013**, *9*, 2339–2354.

[369] M. Krepl, M. Zgarbová, P. Stadlbauer, M. Otyepka, P. Banáš, J. Koča, T. E. Cheatham, P. Jurečka, J. Sponer, *J. Chem. Theory Comput.* **2012**, *8*, 2506–2520.

[370] M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, P. Jurečka, *J. Chem. Theory Comput.* **2015**, *11*, 5723–5736.

[371] B. Hess, H. Bekker, Berendsen, Herman J. C., Fraaije, Johannes G. E. M., *J. Comput. Chem.* **1997**, *18*, 1463–1472.

[372] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577–8593.

[373] Z. Weinberg, R. R. Breaker, *BMC Bioinf.* **2011**, *12*, 3.

[374] L. Bonen, J. Vogel, *Trends Genet.* **2001**, *17*, 322–331.

[375] H. Hwang, S. Myong, *Chem. Soc. Rev.* **2014**, *43*, 1221–1229.

[376] K. V. Morris, J. S. Mattick, *Nat. Rev. Genet.* **2014**, *15*, 423–437.


CMSZH
CHEMICAL MOLECULAR SCIENCES ZURICH
GRADUATE SCHOOL